A Practical Approach to the Analysis of Survival Data.

Thesis

Jennifer Ann Grace Goodwin, M.A. (Oxon.)

A Practical Approach to the Analysis of Survival Data

Dissertation for the degree of Bachelor of Philosophy

Medical Statistics

Submitted May 1975
Awarded Oct. 1975.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

The question of how to compare survival between two or more groups is considered mainly with a view to applications in medical studies. Emphasis is laid on finding methods of comparison which are simple to perform, and on the presentation of the results in a form which may readily be understood by the layman.

The main nonparametric methods of estimating a survival distribution are described, and the connections between them pointed out. Parametric methods considered are those based on the exponential and Weibull distributions. The comparison of two groups is considered in detail, particular attention being paid to simple graphical methods, and to the clarification of some points arising from the recent papers by Cox (1972) and Peto and Peto (1972).

Where account is to be taken of concomitant variables, the important regression models available are those based on the exponential distribution and Cox's model, in which the form of the underlying distribution remains arbitrary. These are presented as extensions of the two-sample models already considered. Where there are many possible covariates, computation becomes heavy and there is a danger that the results may be difficult to interpret. It is suggested that a preliminary analysis of the data in strata by simple two-sample methods be made, as an aid in deciding which covariates to include in the model; a subsequent regression analysis might then be presented as a refinement of this. Such a preliminary

analysis is demonstrated on a body of data from a clinical trial.

0. Introduction

A number of subjects are observed from the occurrence of some initiating event until failure occurs, and their failure times noted. The subjects might be patients after a heart attack, observed from the time of the attack until death, leukaemia patients observed from the beginning of remission until the re-appearance of symptoms, or experimental animals treated with a carcinogen and observed from the start of treatment until the appearance of a tumour. The time until failure is referred to as survival time.

Suppose the subjects may be divided into two or more groups according to some factor which may affect survival. The heart patients, for example, might be given different treatments, the leukaemia patients might be classified as having high or low white blood cell count at diagnosis, and the experimental animals might be treated with two or more different doses of the carcinogen. The object is to estimate the survival distribution for each group and make a comparison.

If numbers are fairly large, it is a simple matter to estimate and draw survival curves (see § 1.1) for the various groups and compare these visually. Where numbers are smaller the analysis is complicated in two ways. The first is that the distribution of failure times is usually far from normal, being considerably skew to the right. This makes conventional methods of analysis difficult to apply, particularly as any difference in shapes between groups is of great importance. The second is the problem

of competing risks. It is seldom that all subjects can be observed until failure. Some of the patients in a clinical trial may withdraw from the trial, or may be still alive or in remission at the end of the trial. Experimental animals may die without developing a tumour. If this happens, we know only that the survival time is greater than some value, called the censoring time. The observation is referred to as right-censored or simply censored.

It is because of these two complications that a number of special techniques have been evolved for dealing with what is otherwise a standard problem in statistics: the estimation and comparison of two or more distributions.

1. Estimating a Survival Distribution

1.1 Theory and notation

Let T be a random variable representing survival time for an individual, and let $F(t)$ be the cumulative distribution function of survival times: $F(t) = pr\{T \leq t\}$. The survivor function is defined as $\mathcal{F}(t) = 1 - F(t) = pr\{T > t\}$, that is, the probability that the individual survives at least time t. The graph of $\mathcal{F}(t)$ is called the survival curve.

$F'(t) = -\mathcal{F}'(t) = f(t)$, the probability density function of survival time.

Because an individual can only fail once, it is convenient to work with the p.d.f. of survival time conditional on failure not having already occurred. This is the age-specific failure rate or hazard $\lambda(t)$. In most medical applications it is sensible to assume that the hazard is a continuous or piecewise continuous function of time. Then

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{pr\{t \leq T < t+\Delta t \mid t \leq T\}}{\Delta t} = \frac{f(t)}{\mathcal{F}(t)} = -\frac{\mathcal{F}'(t)}{\mathcal{F}(t)} \qquad (1.1)$$

$$\lambda(t) = -\frac{d}{dt}\{\ln \mathcal{F}(t)\} \qquad (1.2)$$

$$\therefore \mathcal{F}(t) = \exp\left\{-\int_0^t \lambda(x)\,dx\right\} = \exp\{-H(t)\} \qquad (1.3)$$

where $H(t)$ is the cumulative hazard function.

These relations show that the distribution is uniquely determined by any of the five functions $F(t)$, $f(t)$, $\mathcal{F}(t)$, $\lambda(t)$, $H(t)$. Note that, from (1.2), $-\lambda(t)$

represents the gradient of the log survival curve (the graph of $\ln \mathcal{F}(t) \equiv -H(t)$).

If $\lambda(t)$ is a discrete function of time, failure is only permitted at times t_i for which there is a non-zero hazard. Although not realistic for medical studies, this model is useful for theory when dealing with nonparametric methods. Then $\lambda(t_i) = p_i$, the conditional probability that an individual fails at time t_i, given that he has not already failed. Since it is no longer true that $f(t) = -\mathcal{F}'(t)$, equations (1.2) and (1.3) no longer hold. The cumulative hazard function is now defined by

$$H(t) = \sum_{t_i \leq t} \lambda(t_i) \quad . \tag{1.4}$$

In estimating the survival distribution for a sample, we assume that the failure times for individuals are independently and identically distributed. It is necessary to distinguish between two types of data, grouped and exact. With grouped data the total observation period is divided into time intervals $[\tau_0, \tau_1)$, $[\tau_1, \tau_2), \ldots, [\tau_{k-1}, \tau_k)$, where $\tau_0 = 0$. In the ith interval $[\tau_{i-1}, \tau_i)$ there are r_i subjects at risk at the beginning of the interval, and m_i failures and ℓ_i censorings during the interval. Then $r_{i+1} = r_i - m_i - \ell_i$. Provided that the intervals are fairly short, little information will be lost by assuming that censorings occur at the ends of the intervals, after the failures; in order to simplify the notation, this assumption will be made throughout.

If observation is continuous, so that exact times of

failure or censoring are known, we may write these in order of occurrence as

$$t_1 , t_2 , t_1^* , t_3 , t_2^* , \ldots$$

where an asterisk denotes a censored observation. For truly continuous observation only one failure or censoring may occur at any given time; however, we extend this case to include slight grouping so that m_i failures may be recorded at time t_i, which is still taken to be exact. Just before time t_i there are r_i subjects at risk. If a failure and a censoring are both recorded at t_i it will be assumed that the failure occurs first.

1.2 Some nonparametric methods of estimation

An actuarial method

Actuarial methods are used where numbers are fairly large, and involve some assumption about the form of the hazard. If we assume that the hazard is a constant λ_i over the ith interval, so that the survival curve is a series of connected straight lines, a reasonable estimate for λ_i is the mid-interval failure rate given by

$$(\tau_i - \tau_{i-1})\,\hat{\lambda}_i = \frac{m_i}{r_i - \frac{1}{2} m_i} . \qquad (1.5)$$

Using (1.3), the survivor function is estimated by

$$\hat{\mathcal{F}}(\tau_i) = \prod_{l=1}^{i} e^{-\hat{\lambda}_l} = e^{-\sum_{l=1}^{i} \hat{\lambda}_l}$$

and $\hat{\mathcal{F}}(t)$ may be found for t in any interval by interpolation.

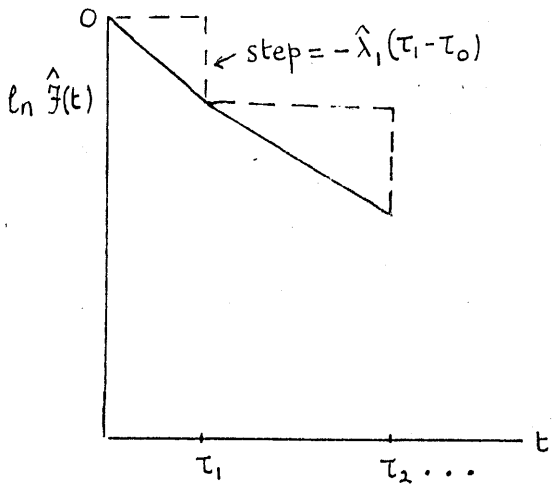

Fig.1 Log survival curve: actuarial estimate.

The life-table method

This method is also used for grouped data. Without making any assumption about the hazard, the probability of failing in the ith interval is estimated by

$$\hat{p}_i = \frac{m_i}{r_i}$$

and if these are regarded as independent probability estimates the survivor function at $\tau_1, \tau_2, \ldots, \tau_k$ is estimated by $\hat{\mathcal{F}}(\tau_i) = \prod_{l=1}^{i} (1-\hat{p}_l)$.

$\hat{\mathcal{F}}(t)$ is undefined elsewhere, but values may be interpolated by imposing some assumption about the hazard.

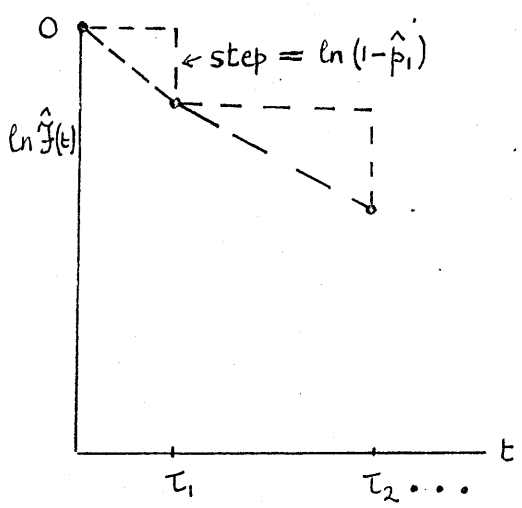

Fig.2 Log survival curve: life-table estimate.

If the hazard is assumed constant over the interval $[\tau_{i-1}, \tau_i)$ its value is estimated from (1.3) by

$$(\tau_i - \tau_{i-1})\hat{\lambda}_i = -\ln\left(1 - \frac{m_i}{r_i}\right) \tag{1.6}$$

Note that this is the same as (1.5), to the second order in $\frac{m_i}{r_i}$.

The product-limit method

Where the sample is small and exact times of failure known, the observation intervals may be made arbitrarily small so that each includes at most one failure. Then the probability of surviving the ith interval is estimated by

$$\hat{p}_i = \begin{cases} 1 - \frac{1}{r_i} & \text{if the interval contains a failure} \\ 1 & \text{otherwise.} \end{cases}$$

Clearly, intervals in which no failure occurs produce no change in the estimated survivor function. Then if the t_i are the recorded failure-times,

$$\hat{F}(t) = \prod_{t_i \le t} \left(1 - \frac{1}{r_i}\right) , \tag{1.7}$$

and if there is slight grouping so that m_i failures are recorded at t_i,

$$\hat{F}(t) = \prod_{t_i \le t} \left(1 - \frac{m_i}{r_i}\right) . \tag{1.8}$$

Kaplan and Meier (1958) showed that the product-limit (PL) estimate given by (1.7) or (1.8) is the maximum

likelihood (ML) estimate in the family of all possible distributions. It is a step-function corresponding to a discrete hazard with values estimated by

$$\hat{\lambda}(t) = \begin{cases} \frac{m_i}{r_i} & t = t_i\,,\ i=1,2,\ldots \\ 0 & \text{elsewhere} \end{cases} \qquad (1.9)$$

so that

$$\hat{H}(t) = \sum_{t_i \leq t} \frac{m_i}{r_i} \qquad (1.10)$$

gives the ML estimate of the cumulative hazard.

Note that, from (1.8) and (1.10), $\ln \hat{\mathcal{F}}(t) \doteq -\hat{H}(t)$, to the first order in $\frac{m_i}{r_i}$, so that the relation of (1.3) holds approximately between the estimates.

If there is no censoring, successive terms of the product cancel out, and $\hat{\mathcal{F}}(t)$ = proportion surviving at time t. $\hat{\mathcal{F}}(t)$ is zero after the last observation if this is a failure, but indeterminate if it is a censoring.

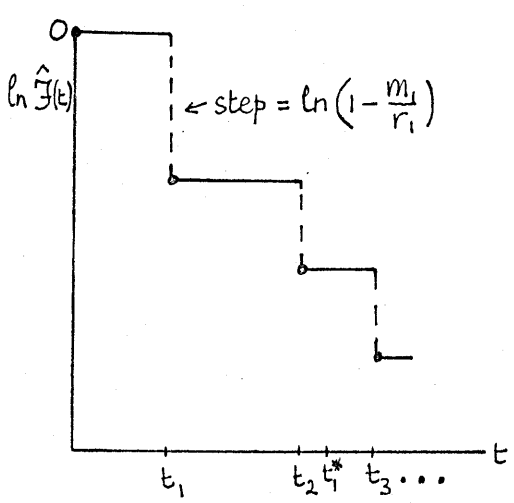

Fig.3 Log survival curve: PL estimate.

Because the discrete-hazard model is unrealistic for medical studies, it is natural to impose some smoothing

on the hazard. This would not necessarily lead to the same ML estimate for $\hat{\mathcal{F}}(t_i)$. However, the values of $\hat{\mathcal{F}}(t_i)$ and $\hat{H}(t_i)$ from (1.8) and (1.10) may be regarded as point-estimates for a model in which the hazard is assumed continuous. Smoothing is then achieved by taking the hazard as constant between failure-times. One such method, used by Breslow, is described in § 2.4.

1.3 Parametric methods of estimation

Distributions specifically proposed for survival problems include the exponential, Weibull and Gompertz. The two former, in particular, often provide a good fit when the failure rate is high, as happens, for example, in clinical trials designed to compare mortality from cancer or heart disease.

The exponential distribution

Suppose the hazard is constant, so that $\lambda(t) = \lambda$ for all t. Then $\mathcal{F}(t) = e^{-\lambda t}$, $f(t) = \lambda e^{-\lambda t}$, and the expected value of survival time = $E(t) = \frac{1}{\lambda}$.

Let N subjects be observed to failure or censoring, and n of these fail. The contribution to the likelihood from a failure at time t_i is $f(t_i) = \lambda e^{-\lambda t_i}$, and that from a censoring at time t_j^* is $\mathcal{F}(t_j^*) = e^{-\lambda t_j^*}$. The log-likelihood is

$$L = -\lambda \sum_j t_j^* + n \ln \lambda - \lambda \sum_i t_i$$

and the ML estimate of λ is given by

$$\hat{\lambda} = \frac{n}{\sum_j t_j^* + \sum_i t_i} = \frac{\text{number of deaths}}{\text{total of observation times}} \quad (1.11)$$

The Weibull distribution

A common form of departure from the exponential model is that $\mathcal{F}(t) = \exp\{-(\lambda t)^\nu\}$, where $\nu \neq 1$. Then $\lambda(t) = \nu\lambda^\nu t^{\nu-1}$, that is, the hazard is monotonic increasing if $\nu > 1$, and monotonic decreasing if $\nu < 1$.

With the notation of 1.1, $pr\{T^\nu > t\} = pr\{T > t^{1/\nu}\} = e^{-\lambda^\nu t}$, so that T^ν is exponentially distributed with parameter λ^ν. Hence if ν is known, the methods outlined for the exponential distribution may be used to find λ. If ν is not known, iterative methods are required for the solution of the ML equations for λ and ν.

Modified exponential and Weibull distributions

For certain animal tumour experiments and observations on times in remission for leukaemia patients, Zelen (1966) suggests the model with survivor function

$$\mathcal{F}(t) = e^{-\lambda(t-\zeta)} \qquad (t > \zeta)$$

in which there is a time-lag before the downward exponential trend begins. Joint ML estimation of λ and ζ is straightforward.

A model which provides a good fit in many carcinogenesis experiments has

$$\mathcal{F}(t) = e^{-\lambda(t-\omega)^k} \qquad (t > \omega).$$

Once ω and k are known, the ML estimate of λ may readily be obtained. However, joint estimation of ω and k is difficult as the distribution is almost degenerate (Peto and Lee, 1973). It is often possible to fix one of these parameters according to its physical meaning.

1.4 Summary

Some terms used in the analysis of survival data are defined. Nonparametric methods of estimating the survivor function from the data are considered, the chief of these being the product-limit method. Parametric methods based on maximum likelihood are outlined for the exponential and Weibull distributions.

2. The Two-sample Case

Very often, survival experience is to be compared for just two groups; some typical examples are given in § 0. For this section we assume that any additional factors, such as age, which might affect survival are similarly distributed over the two groups.

2.1 Parametric methods of comparison

Underlying exponential distribution

If we may assume that the hazards for the two groups are constants λ_1, λ_2 respectively, the method of comparison is particularly straightforward. Let $\mathcal{T}_1$, $\mathcal{T}_2$ be the totals of observation times, and n_1, n_2 the numbers of failures for the two groups. Then, using (1.11), the ratio of the hazards is estimated by

$$\frac{\hat{\lambda}_2}{\hat{\lambda}_1} = \frac{\mathcal{T}_1 n_2}{\mathcal{T}_2 n_1} \quad .$$

It can be shown that $2\lambda_j \mathcal{T}_j$ is distributed as χ^2 with $2n_j$ degrees of freedom ($j = 1,2$), so that

$$\frac{\lambda_2}{\lambda_1} = \frac{\mathcal{T}_1 n_2}{\mathcal{T}_2 n_1} \times F_{(2n_2, 2n_1)} \ ,$$

yielding a test of the null hypothesis $\frac{\lambda_2}{\lambda_1} = 1$, and confidence limits for the ratio of hazards.

Illustration: Table 1 shows a set of data of Freireich et al quoted by Cox (1972), on times of remission of two groups of leukaemia patients.

Table 1. Times of remission of leukaemia patients (weeks)

Group 1 (drug 6-MP): 6*,6,6,6,7,9*,10*,10,11*,13,16,17*, 19*,20*,22,23,25*,32*,32*,34*,35*

Group 2 (control): 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12, 12,15,17,22,23

(* denotes censored observation)

$T_1 = 359,\ n_1 = 9$
$T_2 = 182,\ n_2 = 21$

$$\frac{\hat{\lambda}_2}{\hat{\lambda}_1} = \frac{359 \times 21}{182 \times 9} = 4.6$$

Upper and lower 2½% points of $F_{(42,18)}$ are 2.19, 0.471 respectively

$\therefore$ 95% confidence interval for $\frac{\lambda_2}{\lambda_1}$ is $\left(\frac{\cdot 471 \times 359 \times 21}{182 \times 9}, \frac{2 \cdot 19 \times 359 \times 21}{182 \times 9}\right)$

$= (2 \cdot 5, 11 \cdot 3)$

Underlying Weibull distribution

If a common index ν is assumed for the two curves, the survivor functions may be written as

$$\mathcal{F}_1(t) = \exp\{-(\lambda_1 t)^\nu\}, \quad \mathcal{F}_2(t) = \exp\{-(\lambda_2 t)^\nu\},$$

giving the constant ratio of hazards as $\left(\frac{\lambda_2}{\lambda_1}\right)^\nu$. If $\nu = 1$, the Weibull distribution reduces to the exponential.

For the data of Table 1, Cox (1972) gives the ML estimate of the index as $\hat{\nu} = 1.3$, and finds that this is just significantly different from 1 at the 5% level. Cox's ML estimate of $\frac{\lambda_2}{\lambda_1}$ is 3.7, which leads to the estimate $\left(\frac{\lambda_2}{\lambda_1}\right)^{\hat{\nu}} = 5.6$ for the ratio of hazards.

Hazard plotting

This gives a quick graphical method of comparing two sets of survival data for which the form of the underlying distribution may be assumed.

Suppose first that the hazards are constant (exponential distribution). Then the cumulative hazards for the two groups are given by

$$H_j(t) = \int_0^t \lambda_j(x)\,dx = \lambda_j t \qquad (j=1,2)$$

The graph of $H_j(t)$ against t is a straight line through the origin, whose slope is λ_j. For each group separately, point estimates of cumulative hazard at the times of failure t_i are found using (1.10). The two cumulative hazards are plotted on the same graph, straight lines fitted by eye, and an estimate of the ratio of hazards obtained. A systematic departure from a straight line would lead to questioning of the assumption of constant hazard. However, in fitting a straight line it must be bourne in mind that the estimate $\hat{H}(t_i)$ becomes increasingly unreliable as the risk set dwindles, so less importance should be attached to plotted points further from the origin. This is especially so when the data are few.

If the Weibull distribution with common index is assumed, $F_j(t) = \exp\{-(\lambda_j t)^\nu\}$ $(j = 1,2)$.

From (1.3) $H_j(t) = (\lambda_j t)^\nu$

$$\log H_j(t) = \nu \log t + \nu \log \lambda_j \qquad (2.1)$$

A plot of $\log \hat{H}_j(t)$ against $\log t$ (or $\hat{H}_j(t)$ against

t on double logarithmic paper) gives a pair of parallel straight lines, with slope ν and intercepts $\nu \log\lambda_1$, $\nu \log\lambda_2$.

Nelson (1972) works out appropriate scales for the axes for a number of alternative distributions. Special hazard plotting papers are available which have suitable non-linear scales on one or both axes, such as the double logarithmic paper mentioned for the Weibull distribution. Nelson also points out that hazard plotting is equivalent to probability plotting for the corresponding cumulative distribution function.

Table 2. Calculation of $\hat{H}(t)$ for the data of Table 1.

	t_i	m_i	r_i	m_i/r_i	$\hat{H}(t_i)=\sum \frac{m_\ell}{r_\ell}$
Group 1:	6	3	21	0.1429	0.14
	7	1	17	0.0588	0.20
	10	1	15	0.0667	0.27
	13	1	12	0.0833	0.35
	16	1	11	0.0909	0.44
	22	1	7	0.1429	0.59
	23	1	6	0.1667	0.75
Group 2:	1	2	21	0.0952	0.10
	2	2	19	0.1053	0.20
	3	1	17	0.0588	0.26
	4	2	16	0.1250	0.38
	5	2	14	0.1429	0.53
	8	4	12	0.3333	0.86
	11	2	8	0.2500	1.11
	12	2	6	0.3333	1.44
	15	1	4	0.2500	1.69
	17	1	3	0.3333	2.03
	22	1	2	0.5000	2.53
	23	1	1	1.0000	3.53

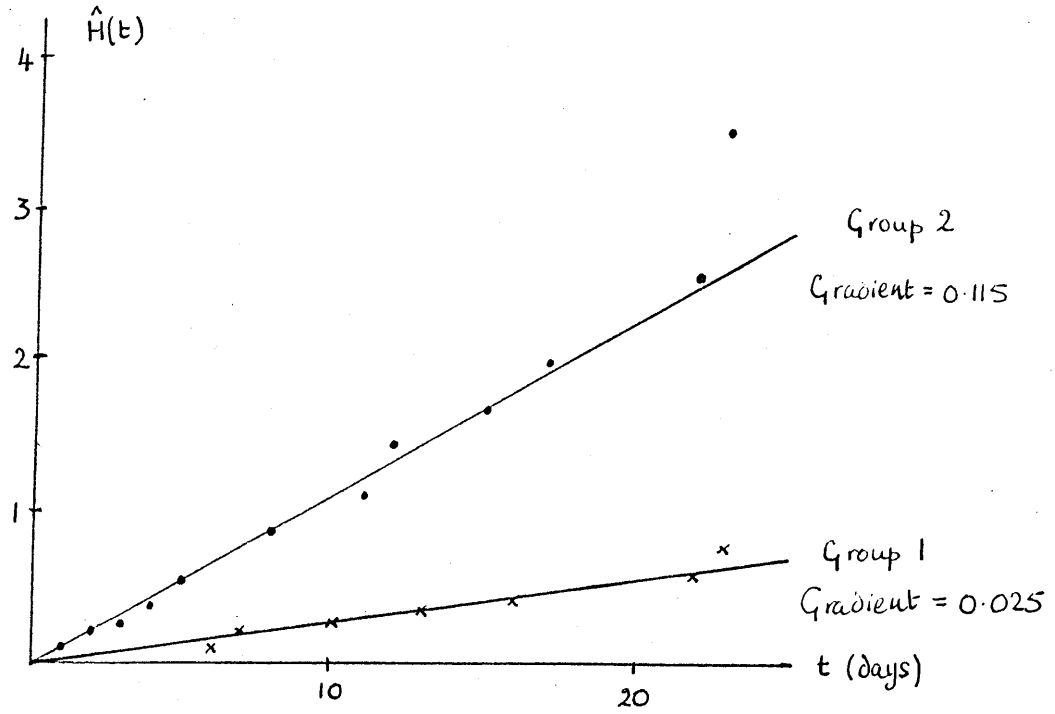

Fig.4 Hazard plot ($\hat{H}(t)$ against t) for the data of Table 1, with straight lines fitted by eye.

Figure 4 shows the hazard plot for the data of Table 1, using linear scales. There is nothing in the diagram to contradict the assumption of an exponential distribution of failure times. From the graph, the estimated ratio of hazards is $\frac{\hat{\lambda}_2}{\hat{\lambda}_1} = \frac{0·115}{0·025} = 4·6$.

In Figure 5 the same hazards are plotted on a log scale (that is, $\log_{10}\hat{H}(t)$ is plotted against $\log_{10} t$ on a linear scale), as for an underlying Weibull distribution. It appears entirely reasonable to fit parallel straight lines, and when this is done the estimated index is found to be 1.2 . It is not possible by this method to estimate the significance of the departure from the value 1.

Taking intercepts from the graph,

$$1·2 \log_{10} \hat{\lambda}_1 \doteq -1·72$$

$$1·2 \log_{10} \hat{\lambda}_2 \doteq -1·08$$

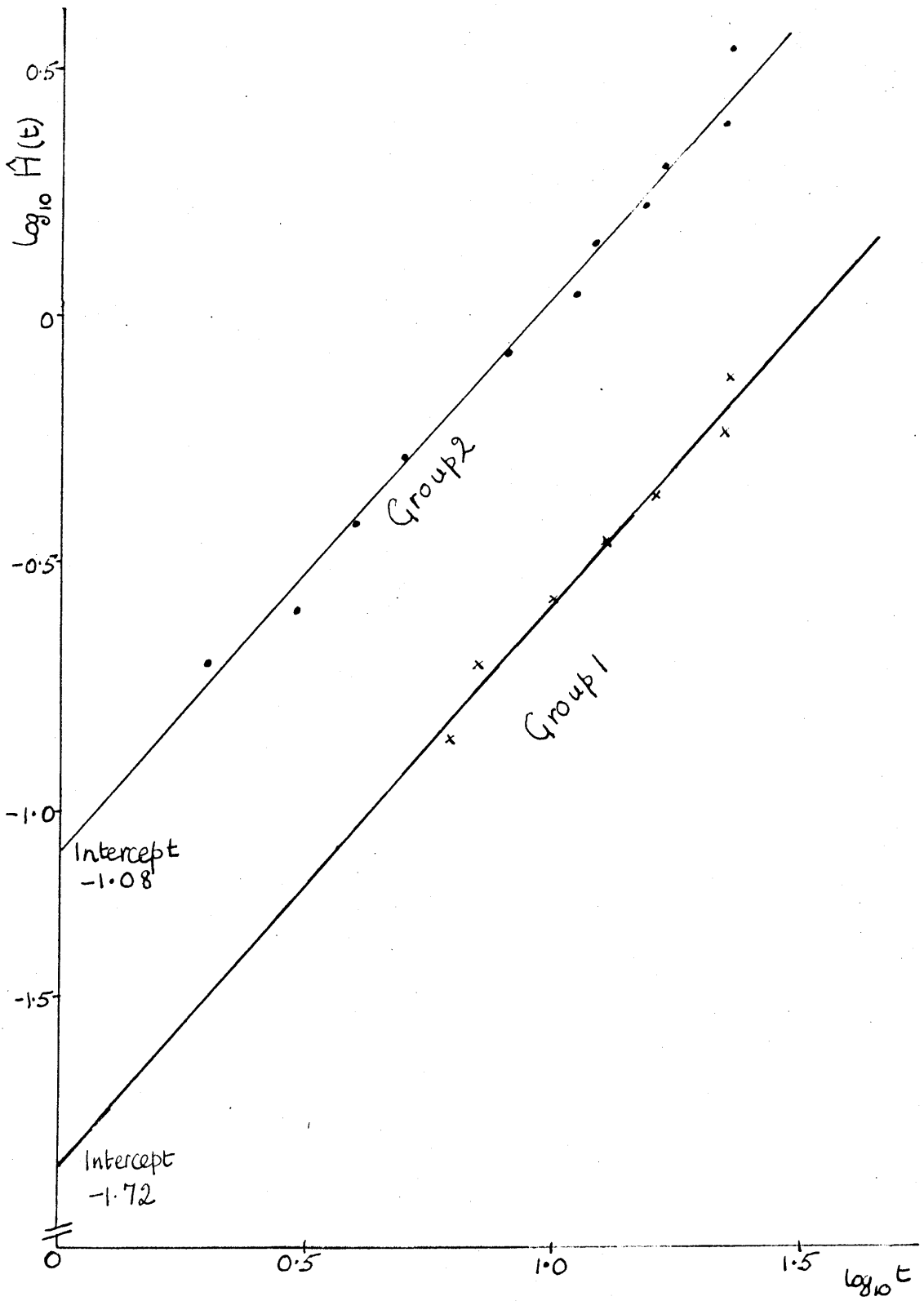

Fig.5 Hazard plot ($\log_{10} \hat{H}(t)$ against $\log_{10} t$) for the data of Table 1, with parallel straight lines fitted by eye.

Hence $1.2 \log_{10}\left(\frac{\hat{\lambda}_2}{\hat{\lambda}_1}\right) \doteq 0.64$, so that $\frac{\hat{\lambda}_2}{\hat{\lambda}_1} = 3.4$, and the estimated ratio of hazards is $(3.4)^{1.2} = 4.4$.

2.2 Simple nonparametric methods of comparison

It is helpful to those who have to interpret the results of a comparative study or an experiment if the comparison can be expressed in terms of a single figure, where it is reasonable to do so. Such figures are also useful for pooling the results of several studies. In the case of the exponential and Weibull distributions, the constant ratio of hazards provides this simple and readily-understood comparison. In this section, we consider three simple methods of comparison which do not rely on any assumption of the precise form of the distribution. One basic condition for the use, on its own, of such a summary is that there should be no crossing of the survival curves.

Comparison of the proportions surviving for a pre-determined period

By the PL method, equation (1.8), the proportions surviving at any pre-determined time t may be calculated for the two samples separately. Note that this method makes use of the survival experience of subjects whose survival times are censored. A variance may be calculated for $\hat{F}_j(t)$ ($j = 1,2$) by taking the logarithm and treating the proportion surviving in each interval as a binomial variable with expected value equal to the observed value. These variances may then be combined

and used in a comparison of the two proportions.

The standardised mortality ratio

Rather than comparing the survival distributions at a single point of time, it seems sensible to use for comparison a single figure which summarises the experience of each group over the whole observation period. Where numbers of subjects are large, one of many possible figures is the standardised mortality ratio (SMR), using the combined sample as standard population.

Let the total observation period be divided into arbitrary intervals as in § 1.2 . For the ith interval we have:

	No. failing	No. surviving	Total
Group 1:	m_{i1}	$r_{i1} - m_{i1}$	r_{i1}
Group 2:	m_{i2}	$r_{i2} - m_{i2}$	r_{i2}
Combined sample:	M_i	N_i	T_i

Then for Group j, according to the experience of the standard population, the expected number of failures is $\frac{M_i}{T_i} \times r_{ij}$. $\text{SMR} = \frac{\text{Observed deaths}}{\text{Expected deaths}}$, so

$$\text{SMR}_1 = \sum_i m_{i1} \Big/ \sum_i \left(\frac{M_i r_{i1}}{T_i}\right) \quad .$$

This is a weighted sum of the ratios $\frac{m_{i1}}{r_{i1}} \Big/ \frac{M_i}{T_i}$ of proportions failing per period, with weights $\frac{M_i r_{i1}}{T_i}$. The comparison of SMRs will therefore be sensible if these ratios are reasonably constant over the total observation period. A preliminary inspection of the ratios should be carried out to make sure that this is the case.

Where numbers are small, such an inspection is unlikely to produce a satisfactory result. However, it will be shown in § 2.3 that the SMR has a particular meaning for survival data which justifies its use, in many cases, even for small samples.

Pooling the separate time intervals

Mantel (1966) points out that a method devised by Mantel and Haenszel (1959) for use in retrospective studies of disease may also be used to compare survival data. Again, the total observation period is divided into arbitrary intervals as in § 1.2, and a contingency table formed for each:

	No. failing	No. surviving	Total
Group 1:	m_{i1}	$r_{i1} - m_{i1}$	r_{i1}
Group 2:	m_{i2}	$r_{i2} - m_{i2}$	r_{i2}
Total:	M_i	N_i	T_i

Given the risk sets and the number dying in the i th interval, under the null hypothesis of a common survival distribution m_{i2} is distributed hypergeometrically with expected value $E(m_{i2}) = \frac{M_i r_{i2}}{T_i}$ and variance $V(m_{i2}) = \frac{r_{i1} r_{i2} M_i N_i}{T_i^2 (T_i - 1)}$. Combining over all the tables, the continuity-corrected statistic

$$\chi^2 = \frac{(|\Sigma m_{i2} - \Sigma E(m_{i2})| - 1/2)^2}{\Sigma V(m_{i2})}$$

is then distributed approximately as χ^2 with one degree of freedom.

If observation is continuous, the time intervals may be made arbitrarily small. Then the tables which

contribute to X^2 correspond to times t_i of single failures, so that $m_{i1}=1$ or 0, $m_{i2}=0$ or 1, $M_i=1$ and $N_i = T_i - 1$ Then $E(m_{i2}) = \frac{r_{i2}}{T_i}$ and $V(m_{i2}) = \frac{r_{i1} r_{i2}}{T_i^2}$.

If p_{ij} $(j = 1,2)$ are the conditional probabilities of failing in the ith interval, the odds ratio $\frac{p_{i2}}{1-p_{i2}} \Big/ \frac{p_{i1}}{1-p_{i1}}$ is estimated by $\frac{m_{i2}}{r_{i2}-m_{i2}} \Big/ \frac{m_{i1}}{r_{i1}-m_{i1}}$. As a summary measure of relative risk, Mantel and Haenszel suggest

$$\mathcal{R} = \frac{\sum m_{i2}(r_{i1} - m_{i1}) / T_i}{\sum m_{i1}(r_{i2} - m_{i2}) / T_i} \quad . \qquad (2.2)$$

This may be expressed as a weighted average of estimated odds ratios, and has the advantage that zero entries present no problem.

For survival data, however, a more suitable measure of relative risk is the ratio of hazards, which is constant for the exponential distribution and for the Weibull distribution with common index. Where numbers of subjects are large this is estimated (§ 1.2) by $\frac{m_{i2}}{r_{i2} - \frac{1}{2}m_{i2}} \Big/ \frac{m_{i1}}{r_{i1} - \frac{1}{2}m_{i1}}$, which suggests that (2.2) might be replaced by

$$\mathcal{R} = \frac{\sum m_{i2}(r_{i1} - \frac{1}{2}m_{i1}) / T_i}{\sum m_{i1}(r_{i2} - \frac{1}{2}m_{i2}) / T_i} \quad . \qquad (2.3)$$

For small numbers and near-continuous observation (§ 1.2), $\frac{m_{i2}}{r_{i2}} \Big/ \frac{m_{i1}}{r_{i1}}$ is the ratio at t_i of the discrete ML estimates of hazard. If this is used, (2.2) becomes

$$\mathcal{R} = \frac{\sum m_{i2} r_{i1} / T_i}{\sum m_{i1} r_{i2} / T_i} \quad . \qquad (2.4)$$

For the data of Table 1, (2.2) gives $\mathcal{R} = 5.2$, while

(2.4) gives $\mathcal{R} = 4.8$.

If only one failure occurs at each t_i , (2.2) and (2.4) are identical, as $m_{i2} = 0$ when $m_{i1} = 1$ and vice-versa.

2.3 Nonparametric models with proportional hazards

By analogy with the exponential and Weibull distributions, we are interested in alternatives to the null hypothesis of a common survival distribution in which the ratio of hazards is constant. By (1.2), the survivor functions may then be expressed as $\mathcal{F}^{\theta_1}(t)$, $\mathcal{F}^{\theta_2}(t)$. Such alternative distributions are known as Lehmann alternatives.

The use of Mantel's procedure against Lehmann alternatives

It may be shown (Radhakrishna, 1965) that the test based on Mantel's χ^2 is highly efficient against alternatives for which the odds ratio (see § 2.2) is constant over all contingency tables. Where numbers are large and each m_{ij} small compared with r_{ij} , the estimated odds ratio $\frac{m_{i2}}{r_{i2}-m_{i2}} \Big/ \frac{m_{i1}}{r_{i1}-m_{i1}}$ is sufficiently close to the estimated ratio of hazards $\frac{m_{i2}}{r_{i2}-\frac{1}{2}m_{i2}} \Big/ \frac{m_{i1}}{r_{i1}-\frac{1}{2}m_{i1}}$ for the test to be efficient against Lehmann alternatives.

For small numbers, it will be shown below in the discussion of Cox's model that the models with proportional odds ratio and proportional hazard are equivalent where observation is continuous. Where there is grouping leading to large numbers of ties this is no longer true, and it is expected that the test based on χ^2 will be relatively inefficient against Lehmann alternatives.

The logrank test of Peto and Peto

Peto and Peto (1972) suggest a test of the null hypothesis of a common survivor function $\mathcal{F}(t)$ against Lehmann alternatives $\mathcal{F}^{\theta_1}(t)$, $\mathcal{F}^{\theta_2}(t)$ for the two groups.

Suppose $\mathcal{F}(t)$ is known and we are interested in finding θ_1. By § 1.1, the contribution to the likelihood from a Group 1 failure at time t is $\theta_1 \mathcal{F}^{(\theta_1-1)}(t)$, and from a censoring at time t^* is $\mathcal{F}^{\theta_1}(t^*)$. The total log likelihood L is a sum of terms $\ln\theta_1+\theta_1\ln\mathcal{F}(t)$, $\theta_1 \ln \mathcal{F}(t^*)$, so that $\frac{dL}{d\theta_1}$ is a sum of terms $\frac{1}{\theta_1} + \ln\mathcal{F}(t)$, $\ln \mathcal{F}(t^*)$.

The ML estimate of θ_1 is found by putting $\frac{dL}{d\theta_1} = 0$. According to the null hypothesis, $E\left(\frac{dL}{d\theta_1}\right)=0$ when $\theta_1 = 1$. If scores $U_i = \left\{\begin{matrix} 1+\ln\mathcal{F}(t_i) \text{ if a failure} \\ \ln\mathcal{F}(t_i) \text{ if a censoring} \end{matrix}\right\}$ occurs at t_i are attached to the observations, the expected value of $X = \sum_{\text{Group 1}} U_i$ is zero under the null hypothesis. Exact significance levels for X may be calculated by randomisation methods, assuming censoring to be independent of group membership.

In practice, $\ln \mathcal{F}(t_i)$ will not be known and must be calculated from the data. Remembering that $\mathcal{F}(t)$ is the common survivor function under the null hypothesis, $\hat{\mathcal{F}}(t_i)$ may be estimated from the combined data by (1.8) and the logarithm taken. Alternatively the first-order approximation, from (1.8),

$$\ln \hat{\mathcal{F}}(t_i) = \sum_{l\le i} \ln\left(1-\frac{m_l}{r_l}\right) \simeq -\sum_{l\le i}\frac{m_l}{r_l}$$

usually eases the computation considerably, and this is the method suggested by Peto and Peto. The quantity $\sum_{l\le i}\frac{m_l}{r_l}$ is the same as the joint estimate $\hat{H}(t_i)$ of cumulative hazard from (1.10). The scores become

$$W_i = \left\{ \begin{array}{ll} 1 - \sum_{\ell \le i} (m_\ell / r_\ell) & \text{if a failure} \\ -\sum_{\ell \le i} (m_\ell / r_\ell) & \text{if a censoring} \end{array} \right\} \text{ occurs at } t_i.$$

Where there are no ties, the W_i depend only on the order in which failures occur, so the test based on them is rank-invariant; it is known as the logrank test. R. Peto (1972) shows that the exact test based on the W_i is of maximal local power among all rank invariant tests, against Lehmann alternatives. It must, however, be emphasised that this locally most powerful test relies on the censoring being independent of group membership, which is seldom the case.

If censoring is not independent of group membership, provided we keep the convention of §1.1 that a censoring cannot coincide with a failure, it is possible to show that $\sum_{\text{Group } j} W_i$ may be treated as a sum of hypergeometric random variables. Peto and Peto suggest a method of calculating the significance level using Pearson curves. Alternatively, they suggest the calculation of a approximate χ^2_1 statistic $\sum_{j=1,2} \frac{(O_j - E_j)^2}{E_j}$, where O_j = observed number of failures, Group j, and E_j = expected number of failures calculated under the null hypothesis: $E_j = \sum_j (-\ln \hat{\mathcal{F}}(t_i))$, where summation is over all times t_i of failures or censorings for Group j. (Note that $\sum_j W_i$ and $O_j - E_j$ are equivalent.) The first of these procedures is time-consuming, the second rather rough.

As an estimator of θ_j, the <u>relative incidence rate</u> for the j th group, Peto and Peto suggest O_j/E_j. That this is a reasonable procedure may be shown as follows:

$$E_j = \sum_j [-\ln \hat{\mathcal{F}}(t_i)] = \frac{1}{\theta_j} \sum_j [-\ln \hat{\mathcal{F}}^{\hat{\theta}_j}(t_i)] \qquad (2.5)$$

Define $d_i = \left\{ \begin{array}{l} 1 \text{ if a failure} \\ 0 \text{ if a censoring} \end{array} \right\}$ occurs at time t_i.

If $\mathcal{F}(t_i)$ is known, the ML estimator of θ_j is

$$\hat{\theta}_j = \frac{\sum_j d_i}{\sum_j [-\ln \mathcal{F}(t_i)]} .$$

Replacing $\ln \mathcal{F}(t_i)$ by its estimator $\ln \hat{\mathcal{F}}(t_i)$ and substituting in (2.5) gives $E_j \doteq \frac{1}{\theta_j} \sum_j d_i$, so $\hat{\theta}_j \doteq \frac{O_j}{E_j}$.

Although calculated differently, this is in fact the SMR for the j th group. To show this, we use the notation of § 2.2 and write $-\ln \hat{\mathcal{F}}(t_i) = \sum_{\ell \le i} \frac{M_\ell}{T_\ell}$,

so that $$E_j = \sum_j \sum_{\ell \le i} \frac{M_\ell}{T_\ell} .$$

Once a typical term $\frac{M_\ell}{T_\ell}$ appears in the summation, it re-appears for every subsequent failure or censoring from Group j ; that is, the number of $\frac{M_\ell}{T_\ell}$ is precisely the number $r_{\ell j}$ of Group j subjects at risk at time t_ℓ. Hence

$$E_j = \sum_\ell \frac{M_\ell r_{\ell j}}{T_\ell}$$

and $$\frac{O_j}{E_j} = \left\{ \sum_\ell m_{\ell j} \right\} \Big/ \left\{ \sum_\ell \frac{M_\ell r_{\ell j}}{T_\ell} \right\} = SMR_j .$$

This shows that the use of the SMR is meaningful even for small samples if there is reason to suppose the hazards proportional. The ratio of hazards is estimated by $\left(\frac{O_2}{E_2}\right) \Big/ \left(\frac{O_1}{E_1}\right)$.

<u>Illustration:</u> The calculations for the W-scores for the data of Table 1 are shown in Tables 3 and 4.

Table 3. Calculation of $\ell n\, \hat{\mathcal{F}}(t)$, under the null hypothesis of a common survivor function, for the data of Table 1.

t_i	m_i	r_i	m_i/r_i	$-\ell n\, \hat{\mathcal{F}}(t)$
1	42	2	0.0476	0.0476
2	40	2	0.0500	0.0976
3	38	1	0.0263	0.1239
4	37	2	0.0541	0.1780
5	35	2	0.0571	0.2351
6	33	3	0.0909	0.3260
7	29	1	0.0345	0.3605
8	28	4	0.1429	0.5034
10	23	1	0.0435	0.5469
11	21	2	0.0952	0.6421
12	18	2	0.1111	0.7532
13	16	1	0.0625	0.8157
15	15	1	0.0667	0.8824
16	14	1	0.0714	0.9538
17	13	1	0.0769	1.0307
22	9	2	0.2222	1.2529
23	7	2	0.2858	1.5387

Table 4. Calculation of the Peto and Peto W-scores for the data of Table 1.

Group 1		Group 2	
t_i	W_i	t_i	W_i
6	1 - 0.3260	1	1 - 0.0476
6	1 - 0.3260	1	1 - 0.0476
6	1 - 0.3260	2	1 - 0.0976
6*	- 0.3260	2	1 - 0.0976
7	1 - 0.3605	3	1 - 0.1239
9*	- 0.5034	4	1 - 0.1780
10	1 - 0.5469	4	1 - 0.1780
10*	- 0.5469	5	1 - 0.2351
11*	- 0.6421	5	1 - 0.2351
13	1 - 0.8157	8	1 - 0.5034
16	1 - 0.9538	8	1 - 0.5034
17*	- 1.0307	8	1 - 0.5034
19*	- 1.0307	8	1 - 0.5034
20*	- 1.0307	11	1 - 0.6421
22	1 - 1.2529	11	1 - 0.6421
23	1 - 1.5387	12	1 - 0.7532
25*	- 1.5387	12	1 - 0.7532
32*	- 1.5387	15	1 - 0.8824
32*	- 1.5387	17	1 - 1.0307
34*	- 1.5387	22	1 - 1.2529
35*	- 1.5387	23	1 - 1.5387
$\sum_{\text{Group1}} W_i$ =	9 -19.2505	$\sum_{\text{Group2}} W_i$ =	21 -10.7494

(* denotes a censored observation)

$O_1 = 9$; $E_1 = 19{\cdot}25$ (Relative incidence rate)$_1$ $= \hat{\theta}_1 \doteq \frac{9}{19{\cdot}25} = 0{\cdot}47$

$O_2 = 21$; $E_2 = 10{\cdot}75$ (Relative incidence rate)$_2$ $= \hat{\theta}_2 \doteq \frac{21}{10{\cdot}75} = 1{\cdot}95$

The ratio of hazards is estimated by $\dfrac{\hat{\theta}_2}{\hat{\theta}_1} \doteq 4{\cdot}2$.

Although shown here for interest, this method of calculation is seldom used in practice, as the W-score is the same as Cox's $U(0)$ (see below), and the use of Cox's formula is computationally simpler.

Cox's model

Cox (1972) proposed a regression model for hazards which will be considered more fully in §3. In the two-sample case, given that observation is continuous, this reduces to

$$\left.\begin{aligned} \lambda_1(t) &= \lambda_0(t) \\ \lambda_2(t) &= e^{\beta}\lambda_0(t). \end{aligned}\right\} \qquad (2.6)$$

where $\lambda_0(t)$ is an arbitrary function of t. This is simply the model with proportional hazards.

It is not possible to obtain the full likelihood for β without specifying $\lambda_0(t)$. Cox obtains a partial likelihood by a somewhat intuitive argument. Given that a failure occurs at t_i, the probability, conditional on the set $\mathcal{R}(t_i)$ of subjects at risk at t_i, that the failure is on the individual as observed is

$$p_i = \begin{cases} \dfrac{1}{r_{i1} + e^{\beta} r_{i2}} & \text{if the failure is from Group 1} \\ \dfrac{e^{\beta}}{r_{i1} + e^{\beta} r_{i2}} & \text{if the failure is from Group 2.} \end{cases}$$

Multiplying the p_i leads to the log-likelihood

$$L(\beta) = n_2\beta - \sum_{i=1}^{N} \ln\left(r_{i1} + e^{\beta} r_{i2}\right) \qquad (2.7)$$

where n_2 is the number of failures from Group 2 and N is the total number of failures.

It is questionable whether the p_i may be regarded as independent, since $\mathcal{R}(t_i)$ depends on the outcome at all previous failure times. However, a rigorous justification of the full regression model version of (2.7) as a marginal likelihood is given by Kalbfleisch and Prentice (1973). Conditions for this justification in the two-sample case are that $\lambda_0(t)$ is never zero over an open interval and β is independent of t.

The derivatives of (2.7) are

$$U(\beta) = \frac{dL(\beta)}{d\beta} = n_2 - \sum_{i=1}^{N} \frac{e^{\beta} r_{i2}}{r_{i1} + e^{\beta} r_{i2}} \qquad (2.8)$$

and

$$-\mathcal{J}(\beta) = \frac{d^2 L(\beta)}{d\beta^2} = -\sum_{i=1}^{N} \frac{e^{\beta} r_{i1} r_{i2}}{(r_{i1} + e^{\beta} r_{i2})^2} \quad . \qquad (2.9)$$

Cox points out that it is not possible to calculate the expected value of (2.9) without making assumptions about the potential censoring times of individuals who were not censored. However, without making such assumptions it is possible to derive an approximate test of the null hypothesis of equal hazards by treating $U(0)$ as asymptotically $N(0, \mathcal{J}(0))$.

If ties are present in the data, the partial likelihood is less easy to obtain. Cox's approach has been much criticised, and several authors have proposed alternatives. Cox treats time as discrete, so that $\lambda(t)$ becomes the conditional probability of failing at time t. As it is now possible for more than one subject to fail at time t_i, the partial likelihood must now be made

conditional also on the number M_i of failures at t_i. Cox replaces (2.6) by

$$\left.\begin{aligned} \lambda_1(t) &= \lambda_0(t) \\ \frac{\lambda_2(t)}{1-\lambda_2(t)} &= e^{\beta}\frac{\lambda_0(t)}{1-\lambda_0(t)} \end{aligned}\right\} \qquad (2.10)$$

which is a model for proportional odds-ratios. The reason for this rather unexpected step is simply convenience. Suppose for example that $r_{i1} = 2$ and $r_{i2} = 3$ Given that one failure occurs at t_i, the probability that this is on the individual as observed is now unpleasantly complicated. However, we can write down the probability that the failure is from Group 1, which is

$$p_i = \frac{2\lambda_1(t_i)[1-\lambda_2(t_i)]}{2\lambda_1(t_i)[1-\lambda_2(t_i)] + 3\lambda_2(t_i)[1-\lambda_1(t_i)]} .$$

Use of (2.10) makes this independent of $\lambda_0(t_i)$, whereas (2.6) does not.

In general, p_i is the probability, given M_i failures at t_i, that, as observed, m_{i1} of these are from Group 1 and m_{i2} from Group 2. Cox's partial likelihood consists of terms

$$\frac{m_{i1} + e^{\beta} m_{i2}}{\sum (m'_{i1} + e^{\beta} m'_{i2})}$$

where summation is over all distinct sets of M_i subjects drawn from the risk set at time t_i, and in any given set m'_{i1} of these are from Group 1 and m'_{i2} from Group 2. This leads to

$$U(0) = n_2 - \sum_{i=1}^{k} M_i A_i \qquad (2.11)$$

and $$\mathcal{J}(0) = \sum_{i=1}^{k} \frac{M_i\,(T_i - M_i)}{T_i - 1} A_i(1-A_i) \qquad (2.12)$$

where $A_i = r_{i2}/T_i$ is the proportion of the risk set in Group 2 and k the number of distinct times of failure.

If $M_i = 1$ for all i, this gives the same result as (2.8) and (2.9).

The test is formally identical with that of Mantel (§2.2), as may be shown by a slight change of notation. (The Cox statistic does not include the continuity correction.) This is not surprising, as both follow from the assumption of proportional odds ratios. Less obvious is the fact that $U(0)$ is identical with the Peto and Peto statistic $\sum_{\text{Group 2}} W_i$. It was shown above that $E_j = \sum_i \frac{M_i r_{ij}}{T_i}$; hence $\sum_{\text{Group 2}} W_i = O_2 - E_2 = n_2 - \sum_i \frac{M_i\, r_{i2}}{T_i} = U(0)$.

If the m_{ij} are all small compared with the r_{ij}, or if the grouping of continuous time is slight, the $\chi(t)$ in (2.10) are small and (2.10) approximates to (2.6). This is not so if there are many ties amongst a small set of data; in this case the model with proportional odds-ratios departs considerably from that with proportional hazards, and one of the alternatives outlined below might be preferred.

Cox demonstrates the calculations for (2.11) and (2.12) on the data of Table 1, obtaining $U(0) = 10.25$ (as we obtained in Table 4 for the sum of the W-scores) and $\mathcal{J}(0) = 6.257$. Estimation of β involves iterative solution of the ML equation. Cox obtains, for these data, $\hat{\beta} = 1.65$, giving $e^{\hat{\beta}}$ = estimated ratio of hazards = 5.21 .

Perhaps not surprisingly, this is close to the pooled estimate of odds ratio from equation (2.2).

Alternatives to Cox's method of dealing with tied data

Kalbfleisch and Prentice (1973) treat tied data as the result of approximations in the measurement of continuous time. If m_i subjects are recorded as failing at time t_i, they assume that these failed in some order unknown because of the grouping of the time-scale. There are $m_i!$ possible orders of failure, each of which contributes to the partial likelihood, which becomes a product of k sums of $m_i!$ terms ($i = 1,2,\ldots,k$). Whilst this approach is intuitively appealing, the likelihood is complicated to write down, and does not lead, in the two-sample case, to a simple formula for variance to replace (2.12). The resulting ML estimate of e^{β} is likely to be closer to the true ratio of hazards, rather than the odds ratio. Kalbfleisch and Prentice illustrate this using grouped simulated data from exponential distributions. For the data of Table 1 they obtain $\hat{\beta} = 1.59$, so their estimate of the ratio of hazards is $e^{\hat{\beta}} = 4.9$.

Breslow (1974) derives a form of the log-likelihood which approximates to those of Cox and of Kalbfleisch and Prentice, but is much simpler. For two groups, the derivation is as follows. From (2.6) and § 1.2, ML estimates of hazard give discrete hazards with values at t_i

$$\lambda_{1i} = \lambda_{oi}, \qquad \lambda_{2i} = e^{\beta}\lambda_{oi}$$

estimated by $\hat{\lambda}_{oi} = \frac{m_{i1}}{r_{i1}}, \qquad e^{\hat{\beta}}\hat{\lambda}_{oi} = \frac{m_{i2}}{r_{i2}}$.

Combining these with weights equal to the numbers at risk leads to the joint estimate

$$\hat{\lambda}_{oi} = \frac{M_i}{r_{i1} + e^{\beta} r_{i2}} \quad . \tag{2.13}$$

Since this also estimates the probability of failing at time t_i, the log-likelihood is

$$L(\beta) = n_2 \beta - \sum_{i=1}^{N} M_i \ln(r_{i1} + e^{\beta} r_{i2}) \quad . \tag{2.14}$$

For testing the null hypothesis of equal hazards, this gives the same statistic $U(0)$ as Cox's method, but with variance

$$\mathcal{J}'(0) = \sum_{i=1}^{N} M_i A_i (1 - A_i) \tag{2.15}$$

This is slightly larger than $\mathcal{J}(0)$ and less likely to lead to a falsely significant result where ties are many.

The likelihoods of Cox, Kalbfleisch and Prentice, and Breslow are identical where there are no ties.

2.4 Further discussion

Estimation of the underlying survivor function

Various methods are suggested by Cox, Kalbfleisch and Prentice, and Breslow for using the combined data to obtain an estimate of the underlying survivor function based on the unspecified hazard $\lambda_o(t)$. Breslow's is the most straightforward. Combining the discrete hazard estimates (2.13) and replacing β by its ML estimate $\hat{\beta}$ gives

$$\hat{\mathcal{J}}(t_i) = \prod_{l=1}^{i} \left(1 - \frac{M_l}{r_{l1} + e^{\hat{\beta}} r_{l2}}\right) \quad . \tag{2.16}$$

The hazard is smoothed by taking it to be constant between failure times. Then, since (2.13) represents the increase in cumulative hazard over $(t_{i-1}, t_i]$, if all censorings in the interval $[t_{i-1}, t_i)$ are adjusted to occur just before t_{i-1}, the estimated hazard is given by

$$(t_i - t_{i-1})\hat{\lambda}_0(t) = \frac{M_i}{r_{i1} + e^{\beta} r_{i2}}, \qquad t_{i-1} < t \leq t_i$$

Values of $\hat{\mathcal{F}}(t_i)$ plotted using (2.16) are then joined by straight lines.

Departures from the model of proportional hazards

If the survival curves cross, so that the short-term prospects for survival are better for one group and the long-term prospects better for the other, it is absurd to try to sum up the differences in survival by a single comparison. Plotting the two estimated survivor functions or cumulative hazards on one graph will reveal any important crossing-over. If the cross-over occurs early in the observation period, it may be caused by some incidental factor such as post-operative deaths. The measurement of survival time might then be started from a later point to allow for this.

Where there is no crossing of the survival curves, two kinds of departure from the model of proportional hazards are common. If a treatment for a disease which causes high mortality is successful, the death rate for the successfully-treated group will decrease with time until it reaches that of the general population. If the treatment is moderately successful, it may cause the

postponement of the deaths of a proportion of patients who would otherwise have died early in the observation period. The death rate for the treated group will initially be lower than that of the control group, but will later increase sharply until it reaches that of the untreated patients. Either of these cases may be treated parametrically by fitting Weibull distributions with different indices, but the amount of computation involved is considerable. Small amounts of data, for which the shape of the survival curve is not well-defined, seldom justify such elaborate treatment.

For a simple comparison of survival between two groups, the choice lies between the exponential model and methods based on the work of Cox, Peto and Peto, and Mantel and Haenszel. If the hazards do not appear to be proportional, they will certainly not both be constant; the test for equality of hazards using Cox's standardised normal deviate seems the most reasonable one to use in this case. The quantity $e^{\hat{\beta}}$ may still be calculated and quoted as an overall measure of group effect; alternatively, the Peto and Peto relative incidence rate (SMR) may be used. The latter is more easily found as iteration is not required, and the necessary calculations are included in the calculations for $U(o)$.

Cox suggests that departures from the model of proportional hazards should then be investigated graphically. The underlying survivor function is estimated; we suggest Breslow's estimate (2.16). Estimates for the two groups, such that the fitted hazards are proportional, are

given by (2.16) for Group 1, and by

$$\hat{\mathcal{F}}_2(t_i) = \left\{\hat{\mathcal{F}}(t_i)\right\}^{e^{\hat{\beta}}}$$

for Group 2. These may be plotted on one graph with the separately-estimated survivor functions, when any marked departure from the model of proportional hazards becomes apparent.

Cox's paper includes such a graph for the data of Table 1, based on $\hat{\lambda}_{oi}$ found by his own iterative method.

A similar graphical investigation is suggested by the form of the model used by Peto and Peto, and the resulting graph is shown in Figure 6 for the data of Table 1. The

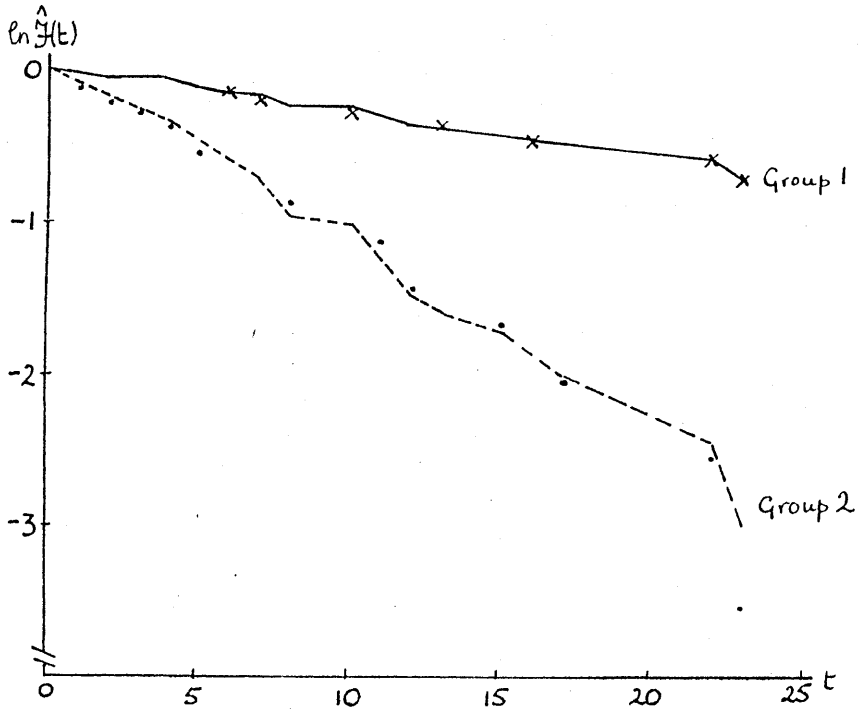

Fig.6 Log survival curves for the data of Table 1.

xx, ·· separate estimates of $\ln \hat{\mathcal{F}}(t_i)$

——, - - - estimates constrained by proportionality

(for clarity, these are joined by straight lines)

combined data provides an estimate of $\ln \hat{\mathcal{F}}(t_i)$, as in Table 3. The relative incidence rates $\hat{\theta}_1$, $\hat{\theta}_2$ for the two groups are then found, and the estimates for the two log survivor functions

$$\ln \hat{\mathcal{F}}_1(t_i) = \hat{\theta}_1 \ln \hat{\mathcal{F}}(t_i), \quad \ln \hat{\mathcal{F}}_2(t_i) = \hat{\theta}_2 \ln \hat{\mathcal{F}}(t_i)$$

are such that the hazards are proportional. This method, whilst it is simpler, is less refined than that suggested by Cox, as the joint estimate of survivor function is based on the null hypothesis of a common hazard for the two groups, rather than the alternative of proportional hazards.

Although the plotting of such graphs may be useful as a clear way of presenting the results, it is seldom essential in order to see whether the hazards are proportional. When the latter is true, we may write the log survivor functions as

$$\ln \mathcal{F}(t), \qquad \theta \ln \mathcal{F}(t) .$$

A plot of the survival curves using a log scale (or, equivalently, a hazard plot) will usually be sufficient to reveal any important difference in shape, especially if one of the hazards is nearly constant.

2.5 Comparison of more than two samples

If the hazards may be assumed constant, the ML and hazard-plotting methods extend straightforwardly to any number of groups. The fitting of Weibull distributions with a common index leads to heavy computation or to the problem of fitting several parallel straight lines by eye

to sets of plotted points.

Extension of Mantel's method of pooling leads to the consideration of an s×2 contingency table for each time-interval, where s is the number of groups. The computation involved is heavy (Mantel and Haenszel, 1959), and if the groups correspond to different levels of a factor, such as age or dose of a carcinogen, a more sensible first approach would be to form 2×2 tables for the two extreme levels and test for significance of the difference between these levels.

The approximate χ^2 test of § 2.3 was originally suggested by Peto and Peto for testing for treatment effect between several groups: $\sum_{j=1}^{s} \{(O_j - E_j)^2 / E_j\}$ may be tested against the χ^2_{s-1} distribution. The relative incidence rates O_j/E_j are calculated just as in the two-sample case, and may be used to pinpoint the treatments responsible for the differences (Roe et al, 1970).

2.6 Summary

The comparison of survival data for two groups of subjects is considered for the case where the groups are similar with respect to all factors other than the distinguishing factor. Numerical and graphical methods are outlined for the exponential and Weibull distributions of failure time; these lead to a simple comparison in the form of the constant ratio of hazards. Simple nonparametric methods described are the comparison of proportions surviving for a pre-determined period, the comparison of standardised mortality ratios, and pooling procedures,

after Mantel, for a set of 2x2 contingency tables, one for each sub-division of the observation period. More sophisticated nonparametric methods, based on a model in which the hazards are unspecified but proportional, are those of Peto and Peto and of Cox. The test statistic proposed by Cox is shown to be identical to those of Mantel and of Peto and Peto, and the relative incidence rate suggested by Peto and Peto as a measure of group effect is shown to be the same as the SMR. Alternatives to Cox's method of dealing with tied data proposed by Kalbfleisch and Prentice and by Breslow are considered, and Breslow's method of estimating the underlying survivor function is outlined. Graphical methods are suggested for examining the estimated survivor functions for departures from the model of proportional hazards. Extension to the comparison of several groups is briefly considered.

The methods are illustrated on a set of data quoted by Cox.

3. The Inclusion of Covariates in the Model

3.1 Underlying distribution exponential

The two-sample case with single covariate

Suppose the subjects in the two groups differ markedly with respect to some variable, such as age, which might affect their survival. Allowance may be made for this by introducing a covariate into the model.

A natural extension of the model with proportional hazards assumes a multiplicative model for the hazards, as suggested by Glasser (1967). Suppose, for the ith subject in Group j ($j = 1,2$), x_{ij} is the difference in his age from the overall mean of the two groups. Let the hazard for this patient be $\lambda_j e^{\beta x_{ij}}$. Then if all the patients have the same age, this reduces to the simple exponential model of § 2.1. If not, the hazards for two subjects in the same group are proportional, the constant of proportionality depending on the difference in their ages.

The contribution to the log-likelihood L of the ith subject in Group j is

$$\ln \lambda_j + \beta x_{ij} - \lambda_j t_{ij} e^{\beta x_{ij}}$$

if the subject fails at time t_{ij}, and

$$-\lambda_j t_{ij} e^{\beta x_{ij}}$$

if the observation is censored at time t_{ij}.

The ML equations are

$$\frac{\partial L}{\partial \lambda_1} = 0; \quad \frac{\partial L}{\partial \lambda_2} = 0; \quad \frac{\partial L}{\partial \beta} = 0.$$

These may be solved iteratively using Newton-Raphson methods. Glasser points out that the estimates of the λ_j may be written explicitly in terms of the estimate of β at each iteration, which leads to some simplification. Convenient initial values are $\hat{\beta}^{(0)} = 0$, $\hat{\lambda}_j^{(0)} = n_j / \sum_i t_{ij}$ (the estimates of λ_j from § 2.1).

An estimate of the covariance matrix for $\hat{\beta}$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ may be obtained by inverting the 3x3 matrix of second derivatives of L. We are interested first of all in whether $\hat{\beta}$ differs significantly from zero. This may be tested by comparing $\frac{\hat{\beta}}{var(\hat{\beta})}$ with the standard normal distribution. Alternatively, the log-likelihoods may be compared for $\beta = \hat{\beta}$ and $\beta = 0$, using the fact that $2\left(L(\hat{\beta}) - L(0)\right)$ is distributed as χ^2 with one degree of freedom.

If $\hat{\beta}$ is found to differ significantly from zero, it is convenient to quote the value of $\hat{\beta}$ together with its variance, and also the <u>age-adjusted hazards</u> $\hat{\lambda}_1, \hat{\lambda}_2$. The initial estimates $\hat{\lambda}_1^{(0)}$, $\hat{\lambda}_2^{(0)}$, ignoring age are also of interest, and may be quoted for comparison.

An alternative model assumes an additive effect for the covariate on the expected survival time of the subject. The hazard λ_j is given by

$$\lambda_j = \left(a_j + b\, x_{ij}\right)^{-1} \qquad (j = 1, 2).$$

Solution is again by ML methods; the models are equivalent to the first order in x_{ij}.

Model allowing for interaction

So far, it has been assumed that the effect of age on survival is the same for the two groups. If it is suspected that this is not so, so that there is interaction between age and group effect, we may fit separately hazards $\lambda_1 e^{\beta_1 x_{i1}}$, $\lambda_2 e^{\beta_2 x_{i2}}$, and test for $\beta_1 = \beta_2$.

Feigl and Zelen (1965) give an account for the alternative model $\lambda_j = (a_j + b_j x_{ij})^{-1}$ in the uncensored case; Zippin and Armitage (1966) extend this to the censored case, for a single group.

More than one covariate

Suppose we wish to compare a number of groups, or two groups for which more than one covariate may be of importance. The multiplicative model extends naturally to one in which the hazard for a subject with vector of covariate values $\underset{\sim}{z}$, where $\underset{\sim}{z}' = (z_1, z_2, \ldots, z_p)$, is

$$\lambda = \lambda_0 e^{\underset{\sim}{z}'\underset{\sim}{\beta}}, \qquad \text{(Model I)}$$

where $\underset{\sim}{\beta}$ is a vector of parameters and $\underset{\sim}{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$.

If $p = 2$, z_1 is a (0,1) variable to indicate group membership (an indicator variable), and z_2 is the deviation from mean age (a baseline variable), this is the same as the model with a single covariate: $\lambda_1 = \lambda_0$, $\lambda_2 = e^{\beta_1}\lambda_0$, and $\beta = \beta_2$. A third covariate may be included to take account of interaction.

The model may again be fitted by Newton-Raphson iteration of the ML equations. Convergence is speeded if

each z_ℓ ($\ell = 1,2,\ldots,p$) is adjusted by subtracting from it the overall mean of the z_ℓ for all subjects. A convenient initial value for $\underset{\sim}{\beta}$ is then $\underset{\sim}{\beta}^{(0)} = \underset{\sim}{0}$. The inverse matrix of second partial derivatives is again used to estimate the covariances.

Selection of the relevant covariates and testing the equality of treatment effects is then achieved by fitting submodels in which one or more covariates are eliminated, or one or more treatment groups pooled, and comparing the resulting log-likelihoods. (Twice the difference in log-likelihoods is distributed as χ^2 with f degrees of freedom, where f is the difference in the number of parameters being estimated.) The decision of which covariate z_ℓ to eliminate at any given stage may be based on a comparison of the estimated parameter β_ℓ with its estimated standard error.

The alternative model may also be extended to more than one covariate as

$$\lambda = \lambda_0 \left(1 + \underset{\sim}{z}'\underset{\sim}{\beta}\right)^{-1} \qquad \text{(Model II)}$$

3.2 Cox's regression model

If it is suspected that the underlying hazard is not constant, we may write, for the hazard for an individual with covariate vector $\underset{\sim}{z}$,

$$\lambda(t; \underset{\sim}{z}) = \lambda_0(t) \times e^{\underset{\sim}{z}'\underset{\sim}{\beta}} \qquad \text{(Model III)}$$

where $\lambda_0(t)$ is an unspecified function of t. This is the full regression model which has (2.6) as a special case

(Cox, 1972).

Where there are no tied data, the partial log-likelihood is

$$L(\underset{\sim}{\beta}) = \sum_{i=1}^{k} \underset{\sim}{z}_i' \underset{\sim}{\beta} - \sum_{i=1}^{k} \ln\Big[\sum_{l \in R(t_i)} \exp(\underset{\sim}{z}_l' \underset{\sim}{\beta})\Big] \qquad (3.1)$$

where $R(t_i)$ is the set of individuals at risk at time t_i. Cox obtains this by the argument outlined in § 2.3, and the justification by Kalbfleisch and Prentice (1973) mentioned there holds for (3.1) provided that $\lambda_0(t)$ is never zero over an open interval, and all the z_l are independent of time.

If ties occur among the failure-times, Cox's discrete-time model (2.10), in which $\lambda(t)$ is now the probability of failing at time t, becomes

$$\frac{\lambda(t)}{1-\lambda(t)} = e^{\underset{\sim}{z}'\underset{\sim}{\beta}} \frac{\lambda_0(t)}{1-\lambda_0(t)} \qquad (3.2)$$

leading to the partial log-likelihood

$$L(\underset{\sim}{\beta}) = \sum_{i=1}^{k} \underset{\sim}{s}_i' \underset{\sim}{\beta} - \sum_{i=1}^{k} \ln\Big[\sum_{l \in R(t_i; M_i)} \exp(\underset{\sim}{z}_l \underset{\sim}{\beta})\Big] \qquad (3.3)$$

where $\underset{\sim}{s}_i$ is the sum of $\underset{\sim}{z}$ over individuals failing at t_i and the notation means that the sum is taken over all distinct sets of M_i individuals drawn from $R(t_i)$.

The procedure for selection of relevant covariates is similar to that for the exponential model (§ 3.1).

As in the two-sample case (§ 2.3), there are difficulties with the model where many ties are present

in the data. The alternative treatments for ties proposed by Kalbfleisch and Prentice and by Breslow were outlined in § 2.3 for the simpler case. In particular, for Breslow's treatment, (2.13) extends to

$$\hat{\lambda}_{oi} = \frac{M_i}{\sum_{\ell \in R(t_i)} \exp(z_\ell' \beta)} \tag{3.4}$$

leading to the log-likelihood

$$L(\beta) = \sum_{i=1}^{k} \left[s_i' \beta - M_i \ln \sum_{\ell \in R(t_i)} \exp(z_\ell' \beta) \right] \tag{3.5}$$

Again, Breslow's method for estimating the underlying survivor function is the most straightforward, (2.16) extending to

$$\hat{F}(t_i) = \prod_{\ell=1}^{i} \left(1 - \frac{M_i}{\sum_{\ell \in R(t_i)} \exp(z_\ell' \hat{\beta})} \right). \tag{3.6}$$

3.4 Comparison of the three regression models

Breslow (1974) compares the three models for data from a clinical trial of maintenance therapy for childhood leukaemia. He finds little difference in the results for models I and II. Model III appears to be more powerful in this case for distinguishing the important covariates. This is explained when the underlying survivor function is plotted and found to depart from the exponential form over the early part of the observation time. Breslow concludes that the use of Model III is worth the slightly heavier computation involved.

Breslow also compares estimates of the underlying survivor function calculated from Cox's original model (3.2), from (3.6), and by the actuarial method of § 1.2 . He finds that for practical purposes there is little difference between these for the first sixty or seventy per cent of the observations. This suggests that an actuarial or PL estimate might be made and plotted at the beginning of the analysis as an aid in deciding between the exponential and nonparametric models.

The Cox model might prove especially useful as a substitute for the modified Weibull model for carcinogenesis (§ 1.3). Since the survivor function proposed in the latter model is difficult to fit, it seems sensible instead to use a method which is independent of the precise form of the hazard function.

Cox's paper suggests that time-dependent covariates might be included in his model. This would be useful as the effect of a successful treatment tends to increase with time (§ 2.4). It is, however, difficult to justify; the marginal likelihood obtained by Kalbfleisch and Prentice is valid only when all covariates are independent of time. No such objection applies in the case of the exponential model.

3.5 Summary

The inclusion of a single covariate in the two-sample exponential model is considered in some detail. The model is then extended to include any number of covariates. Cox's regression model, in which the form

of the underlying hazard remains arbitrary, is presented as a similar extension of the model with proportional hazards of § 2.3 . Reference is made to Breslow's comparison of these models.

4. Analysis of a Body of Data

Table 5 gives the available data on 93 patients included in a clinical trial conducted by the Brain Tumour Study Group, M.D.Anderson Hospital and Tumour Institute, University of Texas. Failure times (in this case, times of death) after the start of treatment are given in days. There are few tied failure times, with no multiplicity greater than 3. The failure times of 3 patients were censored; these patients were still alive at the end of the trial. Treatments were coded as follows:

Chemotherapy	X-rays
0: control	0: none
1: treatment	1: between 0 and 3000 rads
	2: between 3000 and 5000 rads
	3: more than 5000 rads

Additional information on each patient consists of duration of symptoms, presumably before the start of treatment, in weeks, age in years, sex, and location of tumour. The latter is coded:

1: frontal
2: temporal
3: parictal
4: occipital
5: deep BG thalamic
6: other

Table 5. Data on 93 patients included in a clinical trial.

Patient Number	Failure Time (days)	Chemo-therapy	X-rays	Duration of Symptoms (weeks)	Age (years)	Sex	Location of Tumour
1	10	0	0	7	46	M	2
2	12	0	2	56	56	F	3
3	15	1	1	9	57	F	1
4	18	1	1	3	39	M	3
5	20	1	0	9	60	M	1
6	21	0	0	24	52	M	6
7	22	1	1	32	60	F	1
8	25	0	0	50	53	M	1
9	30	1	0	14	54	M	6
10	32	1	2	8	57	M	1
11	34	0	0	13	72	M	3
12	37	0	0	19	55	M	3
13	41	1	0	27	67	M	1
14	42	0	0	73	40	M	5
15	46	1	0	14	52	M	2
16	49	0	0	8	57	F	1
17	51	0	3	76	60	M	5
18	51	1	0	60	56	M	1
19	54	0	0	10	40	M	5
20	56	1	3	37	68	F	1
21	57	1	3	14	59	F	3
22	59	1	0	15	36	M	1
23	62	1	2	21	46	F	2
24	64	1	0	19	56	M	3
25	71	1	1	22	60	M	1
26	72	1	2	13	53	F	5
27	79	1	0	15	40	M	4
28	82	1	1	25	45	M	3
29	85	1	0	22	43	F	2
30	97	0	0	23	48	F	1
31	102	0	1	47	52	M	4
32	107	0	0	43	71	M	3
33	108	0	0	18	50	M	3
34	119	0	0	187	48	M	1
35	121	0	0	23	57	M	1
36	129	1	3	24	59	M	2
37	131	1	0	19	59	M	1
38	132	0	0	42	49	M	3
39	134	0	2	44	60	M	3
40	135	1	0	82	35	F	2
41	135	0	3	32	40	M	6
42	136	1	0	31	22	F	3
43	143	1	0	28	52	M	3
44	144	1	0	22	41	M	2
45	145	1	3	23	55	M	2

Table 5 (continued)

Patient Number	Failure Time (days)	Chemo-therapy	X-rays	Duration of Symptoms (weeks)	Age (years)	Sex	Location of Tumour
46	147	1	0	23	45	M	4
47	162	0	0	26	66	F	2
48	162	1	0	30	64	F	4
49	162	1	3	37	50	M	1
50	164	1	0	78	53	F	2
51	177	1	0	49	48	M	2
52	181	0	1	41	45	M	1
53	194	0	0	347	57	M	2
54	200	0	3	312	42	M	2
55	204	1	3	43	57	M	2
56	214	1	2	39	42	M	1
57	231	0	3	38	53	M	1
58	234	0	2	40	60	F	3
59	248	1	3	39	51	F	3
60	252	0	3	39	31	M	2
61	253	0	3	48	47	M	2
62	253	1	3	62	73	M	6
63	255	1	3	40	53	M	3
64	255	0	2	48	70	M	2
65	259	0	0	99	44	M	1
66	264	1	2	42	53	M	1
67	272	1	0	41	55	F	2
68	272	0	3	41	30	M	4
69	274	1	2	42	47	M	2
70	275	1	0	44	56	F	3
71	281	1	2	43	66	M	1
72	297	0	3	50	54	M	2
73	298	0	0	49	51	M	3
74	325	0	3	59	56	M	2
75	336	1	3	153	58	M	1
76	345	1	3	76	52	M	2
77	347	1	2	80	57	M	1
78	359	0	2	57	55	M	1
79	385	0	3	59	59	M	2
80	387	0	3	54	45	F	6
81	408	0	3	85	36	M	3
82	410	1	2	14	68	M	1
83	449	0	3	71	58	M	4
84	466	1	3	140	40	F	2
85	475	0	2	71	57	M	4
86	484	0	0	84	50	M	1
87	495*	0	1	90	53	F	2
88	522	1	3	96	52	M	1
89	526	0	2	87	59	M	2
90	669	1	2	121	47	M	2
91	815*	1	3	168	22	M	1
92	847*	1	3	400	48	F	2
93	1760	1	3	253	27	M	1

(* denotes a censored observation)

There was no random allocation of treatments to patients so chemotherapy and X-rays are to be treated as factors in the same way as age, duration, etc. The problem is to disentangle the effects of the factors on survival.

The use of Cox's regression model is suitable for data of this kind, since there are several factors of interest, and the data are too few for a satisfactory breakdown into homogeneous subgroups. However, the computation involved in applying Cox's method with a large number of covariates is very heavy. Our purpose here is to demonstrate the methods of §2 in the kind of analysis which might usefully be done before embarking on the use of the full regression model, in order to suggest which factors should be included.

4.1 The overall survival curve and marginal distributions of survival time

As there are few censorings, the PL method (§1.2) is the simplest way of estimating $\hat{\mathcal{F}}(t_i)$, which is equal to the proportion surviving for most t_i. Figure 7 shows these values for the whole data, plotted on a log scale; points are plotted for the last two censored observations as a reminder that $\hat{\mathcal{F}}(t)$ does not become zero within the time-scale used. The exponential curve $\hat{\mathcal{F}}(t) = e^{-\hat{\lambda}t}$, where $\hat{\lambda}$ is found using equation (1.11), is also shown; the fit is quite good, although the hazard does appear to be increasing slightly.

For Figure 8, subjects are grouped according to the

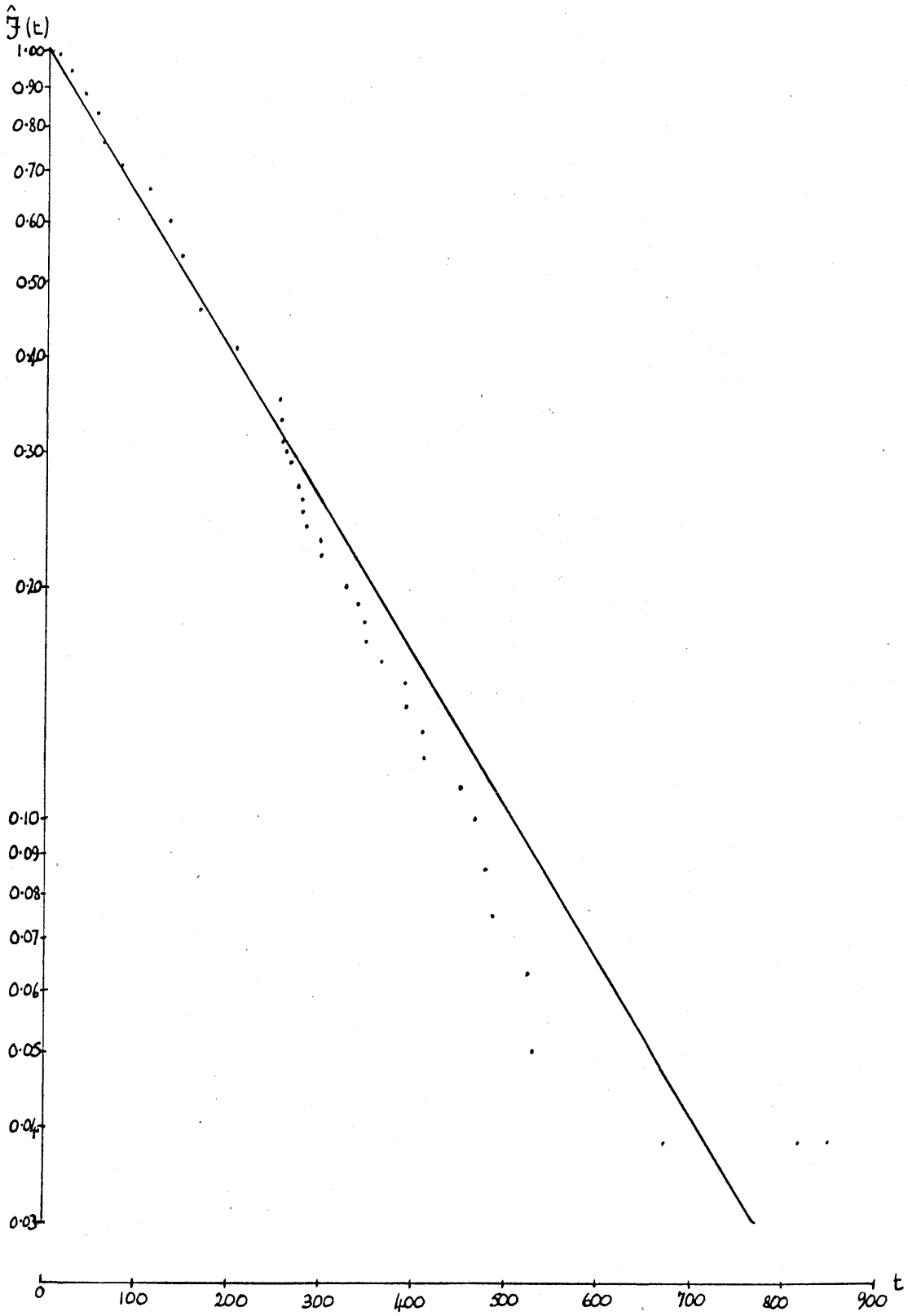

Fig.7 Graph of $\hat{\mathcal{F}}(t)$ (log scale) against t for the data of Table 5 (plotted for every 5th death for $t<250$) with best-fitting exponential curve ($\hat{\lambda} = 0.0044$).

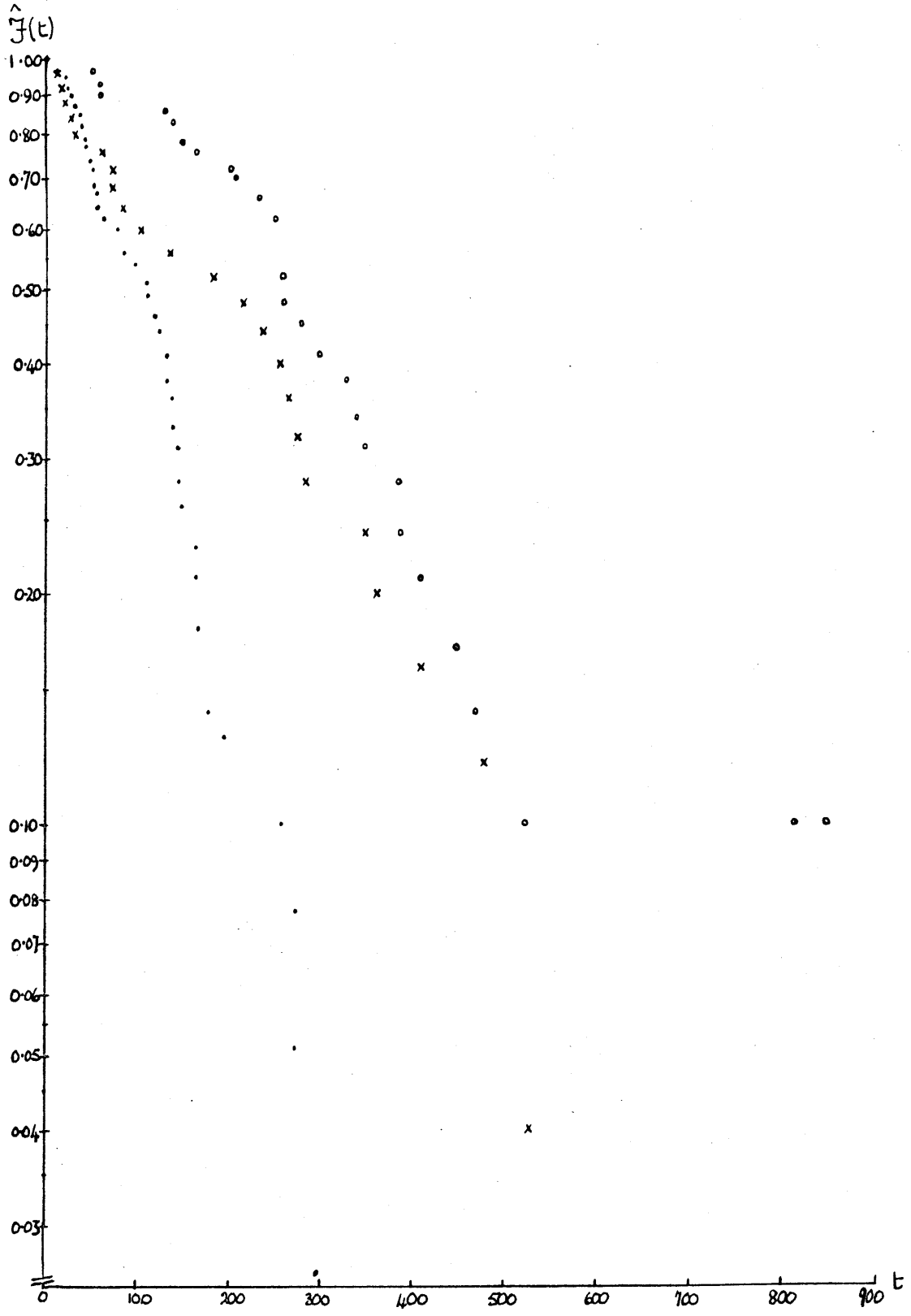

Fig.8 Graph of $\hat{\mathcal{F}}(t)$ (log scale) against t for X-rays 0, 1-2, 3.

·· X-rays 0 ×× X-rays 1-2 ◦◦ X-rays 3

amount of X-rays received. X-rays 1 and 2 are pooled because only 8 patients received X-rays 1. For the most part the curves lie above one another, but there is a marked departure from a straight line for X-rays 3. This suggests that non-parametric methods might be preferable to fitting an exponential model when comparing subgroups of the data.

For the purposes of a preliminary analysis, duration and age need to be broadly grouped. After examining the distributions of these, it was decided to group as follows:

Duration (weeks)	0-29 (Low, L)	30-69 (Medium, M)	70 and over (High, H)
No. of patients	33	37	23

Age (years)	20-49 (Low, L)	50 and over (High, H)
No. of patients	33	60

Only Locations 1, 2, and 3 included sufficient patients (30, 28 and 19, respectively) for a separate comparison between these locations to be possible.

Marginal distributions of survival were examined by calculating Cox's $U(0)$ (2.11) and $J(0)$ (2.12) for two suitably-chosen groups, and combining these to obtain a continuity-corrected standardised normal deviate (SND). The results are shown in Table 6, together with the Peto and Peto relative incidence rates O/E for the two groups in each case.

The important factors appear to be X-rays and

Table 6. Calculation of standardised normal deviates and relative incidence rates for marginal distributions.

Factor	Levels for two groups	Number of deaths	Expected no. deaths, Gp 2 ($E_2 = \sum M_i A_i$)	Variance $\mathcal{J}(o)$	SND	$\frac{O}{E}$
Chemo-therapy	0	41				1.00
	1	49	49.0549	21.5468	0.09	1.00
X-rays	0	39				1.86
	1,2,3	51	68.9894	14.5196	-4.72	0.74
Duration	L	33				2.15
	H	20	37.6294	8.5839	-5.85	0.53
Age	L	31				0.87
	H	59	54.4762	20.2551	1.01	1.08
Sex	M	69				0.99
	F	21	20.0887	15.4328	0.10	1.05
Location	1	29				1.11
	2	26	28.9326	13.3435	-0.67	0.90
Location	1	29				0.87
	3	19	14.5671	9.4205	1.25	1.30

duration of symptoms, but it is clear at a glance from Table 5 that these are related. Although no other factor appears to have a significant effect, it is possible that dependencies between the factors are obscuring such effects. For example, since high duration of symptoms apparently confers a survival advantage, the fact that few older patients have high duration might obscure an advantage due to age. The next step is to examine two-way distributions of all pairs of factors; this reveals a number of possibly important dependencies.

4.2 Two-way distributions excluding survival

Sex: All factors except location and chemotherapy are distributed independently of sex, and in view of the non-significant result in Table 6 it was decided to pool the sexes when considering all other factors.

Age: All factors except duration of symptoms appear to be independent of age. Table 7 shows the age×duration distribution. There is a lack of high durations amongst the older patients; possibly such patients have died without being presented for treatment.

Table 7. Age×Duration

Duration	L	M	H	Total
Age L	11	11	11	33
Age H	22	26	12	70
Total	33	37	23	93

Location: Both chemotherapy and X-rays are unbalanced over locations. There is also a lack of high durations for Location 3, as shown in Table 8.

Table 8. Location×Duration

Duration	L	M	H	Total
Location 1	11	11	8	30
Location 2	8	10	10	28
Location 3	8	10	1	19
Location 4,5,6	6	6	4	16
Total	33	37	23	93

Duration: The lack of independence of age and duration has already been mentioned. Because of the lack of randomisation to treatments, a high proportion of patients with low duration received chemotherapy, and a similarly high proportion were not treated with X-rays.

About 40% of patients with Durations M and H received the highest dose of X-rays, as opposed to roughly 10% of patients with Duration L.

Table 9. Chemotherapy×Duration

Duration		L	M	H	Total
Chemotherapy	0	10	20	12	42
	1	23	17	11	51
Total		33	37	23	93

Table 10. X-rays×Duration

Duration		L	M	H	Total
X-rays	0	22	10	7	39
	1,2	8	12	5	25
	3	3	15	11	29
Total		33	37	23	93

Treatments: Although treatments were clearly non-randomised, it is difficult to see any reason for the way in which they were allocated; their distribution over the other factors appears haphazard. Chemotherapy and X-rays do, however, appear to be independent of oneanother.

4.3 Comparisons of survival for pairs of factors within different cells

Important dependencies are all between duration and another factor. The effects of duration and X-rays are confused, and effects of age, location and chemotherapy may be obscured due to the lack of independence between duration and each of these factors. In an attempt to disentangle the effects of the other factors from that of duration, comparisons of survival are now made within cells of the two-way tables of §4.2.

Once again, the Cox SNDs are calculated as a test of the equality of survival probability between two groups. The relative incidence rate is again given as an indication of the size of an effect, although in some cases numbers are so small that this is only a very rough guide.

X-rays and Duration: It is evident from Table 6 that these are the main factors affecting survival, and Table 10 confirms the impression, gained from Table 5, that they are related. As neither is very closely related to any other factor, we may ignore the other factors when assessing the effects of X-rays and duration.

Numbers are such that we may reasonably compare the effect of X-rays versus no X-rays for pooled Durations M and H, but the comparison must be made separately for Duration L. The results are shown in Table 11. The effect of X-rays appears to be significant for Durations M and H, but not for Duration L. However, this apparent interaction might be due to the higher proportion of patients with durations M and H who received X-rays 3.

Table 11. SNDs and relative incidence rates for X-ray contrasts.

Duration	X-rays	Number of deaths	$E_2 = \sum M_i A_i$	$\vartheta(0)$	SND	O/E
L	0	22				1.05
	1,2,3	11	12.0135	7.1310	-0.19	0.92
M and H	0	17				1.96
	1,2,3	40	48.3301	6.9431	-2.97	0.83

Any attempt to refine this crude analysis further introduces a common problem in statistics. In any further

sub-division, the numbers in the groups become very small. Tests based on them will be unreliable, and examining all possible contrasts greatly increases the chance of obtaining a falsely significant result. It is at this stage that a regression model would provide the best method of proceeding with the analysis.

Comparisons of duration effect were made between Duration L and pooled Durations M and H for the groups with X-rays 0 and with X-rays 1 and 2 respectively. Numbers do not permit such a comparison to be made for X-rays 3. The results are shown in Table 12 and indicate a comparable duration effect for both X-ray groups. Further breakdown into subgroups raises the problems already mentioned in connection with X-rays.

Table 12. SNDs and relative incidence rates for duration contrasts.

X-rays	Duration	Number of deaths	$E_\lambda = \sum M_i A_i$	$\hat{f}(0)$	SND	O/E
0	L	22				1.81
	M and H	17	26.8734	6.8352	-3.59	0.63
1,2	L	8				2.25
	M and H	16	20.4415	2.7870	-2.36	0.78

The shapes of the survival curves for the various subgroups are interesting, and four of these are shown in Figure 9. Here $\hat{\mathcal{F}}(t_i)$ (PL estimate) is plotted, for Durations M and H separately, for the highest and lowest X-ray groups. With X-rays 0, there is no clear distinction between the curves for the two durations. The curve for X-rays 3, Duration M shows an "elbow"

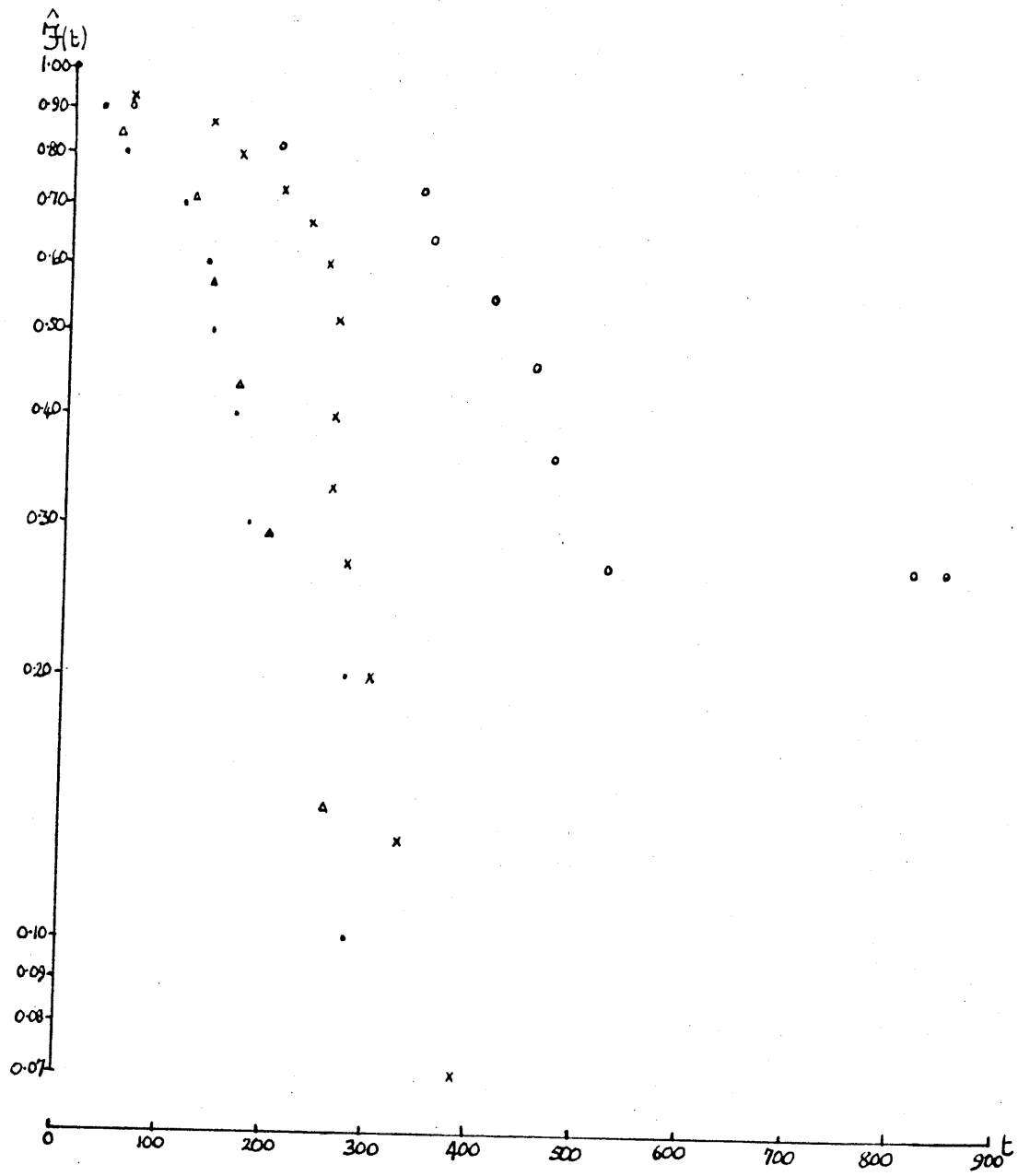

Fig.9 Graph of $\hat{\mathcal{F}}(t)$ (log scale) against t for X-rays 0 and 3, each with Duration M and H.

·· X-rays 0, Duration M
▵▵ X-rays 0, Duration H
×× X-rays 3, Duration M
°° X-rays 3, Duration H

followed by a sharp decline, which suggests an initial advantage for patients receiving this treatment. For Duration H, the curves for the two X-ray levels are well-separated. Although any conclusions from such small numbers must be tentative, the graph does appear to support the impression of an interaction between X-rays and duration, and suggests that X-rays may have a long-term effect only for patients with high duration of symptoms.

Age: The numbers in Table 7 indicate that the survival distributions for age may be compared for pooled Durations L and M, but a separate comparison is necessary for Duration H. For each of these pairs of groups, a check was made of the two-way distribution with chemotherapy, X-rays and location. Only in the last of these was any marked dependency found: for Durations L and M, a majority of the tumours at Locations 1 and 3 are found among the older group of patients.

Results of the comparison are shown in Table 13. No effect of age on survival is apparent.

Table 13. SNDs and relative incidence rates for age contrasts

Duration	Age	Number of deaths	$E_2=\sum M_i A_i$	$\vartheta(0)$	SND	O_i/E_i
M,L	L	22				1.06
	H	48	49.2222	14.0998	-0.20	0.98
H	L	9				0.83
	H	11	9.2117	4.3408	0.62	1.19

Chemotherapy: The proportions of patients with Durations M and H receiving chemotherapy are similar (Table 9), so it is reasonable to pool these durations for comparison:

Duration		L	M and H	Total
Chemotherapy	0	10 (Gp A)	32 (Gp C)	42
	1	23 (Gp B)	28 (Gp D)	51
Total		33	60	93

Before comparing A with B and C with D, checks were made on their two-way distributions with other factors. It was found that no members of Group A received any X-rays, so Group B patients receiving X-rays were ignored in making the comparison, leaving 12 Group B patients. Locations were independently distributed over the resulting groups. X-rays were independently distributed over Groups C and D, but D showed some lack of Locations 3 and 4,5,6.

The results of the comparisons are shown in Table 14. There is no evidence that chemotherapy has any effect on survival, although it is impossible to disentangle the effects of chemotherapy and location because of the very small numbers in comparable groups.

Table 14. SNDs and relative incidence rates for chemotherapy contrasts.

Duration	Chemo-therapy	Number of deaths	$E_2 = \sum M_i A_i$	$\hat{f}(0)$	SND	O/E
L, with X-rays 0	0(A)	10				1.08
	1(B)	12	12.7146	4.9066	-0.10	0.94
M,H	0(C)	31				1.13
	1(D)	26	29.6451	13.3893	-0.86	0.88

Location: Table 6 shows no significant difference between survival distributions for Locations 1 and 2. As two-way tables for location with each of the other factors show them all reasonably independent of Locations 1 and 2, it is unlikely that there is any real difference in survival between the two locations.

Although the numbers in Table 8 suggest that it would be reasonable to pool over Locations 1 and 2 for comparison with Location 3 for Durations L and M, it was decided to compare Location 3 with Location 1 only as this gives independence of X-ray treatments.

From the results in Table 15, no difference in survival between Locations 1 and 3 is apparent for Durations M and L. It is obviously impossible (Table 8) to make any such comparison for Duration H, and investigation of the other locations is also impossible because of the small numbers.

Table 15. SND and relative incidence rate for location contrast.

Duration	Location	Number of deaths	$E_2 = \sum M_i A_i$	$\delta(0)$	SND	O/E
L,M	1	22	17.5316	9.3072	0.01	0.98
	3	18				1.03

4.4 Conclusion

Although the various factors are very confused with one another, it seems unlikely that any have an effect on survival apart from duration of symptoms and X-rays. The effect of duration is the most marked; in general, patients

with a higher duration of symptoms survive for longer, although there is probably some interaction with X-rays. It is possible that X-rays have a long-term effect only on patients with high duration of symptoms.

The data might usefully be investigated more fully using Cox's regression model; this would enable account to be taken of individual lengths of duration of symptoms, instead of the rather arbitrary grouping used. Covariates should be included for X-rays, duration of symptoms, and X-raysxduration interaction.

4.5 Discussion

Although no analysis of a small amount of data can be entirely satisfactory, a regression model does provide a way of making the most of the available information. However, where the number of possible covariates is large, regression methods can become clumsy and tedious to apply. Some preliminary analysis of the data in strata, as demonstrated in this section, becomes essential in order to decide which covariates should be included. Such an analysis also has the advantage of promoting a greater awareness of the ways in which the various factors are related. It is more readily understood by the layman than a full regression analysis, and any fuller investigation using a regression model may then be presented as a refinement of the preliminary analysis.

Two further points should be made in connection with the regression model. The first is that in general, where regression methods are used, some check on the

validity of the model may be made by examining residuals. No such check is available with Cox's model, since the form of the underlying distribution is unspecified. The second point concerns interactions. With the grouping used on the current set of data, 3 levels for X-rays and 3 for duration provide 2x2 = 4 degrees of freedom for interaction between these two factors. This is too many for an intelligible analysis, and by pooling the sub-groups in Table 10 we have reduced the number of degrees of freedom to 1. However, if numbers permitted, we might usefully pool to form a 2x3 table and retain 2 degrees of freedom in order to describe the interaction more fully. In the simplest form of regression model including an interaction term, in which $\underset{\sim}{z}'\underset{\sim}{\beta}$ may be written $\beta_1 z_1 + \beta_2 z_2 + \beta_{12} z_1 z_2$, only one degree of freedom is available for interaction; it is assumed that the effect of X-rays on survival varies linearly over durations. If, as our graphical investigation suggests, this is not the case, X-rays having no effect except where duration of symptoms is high, it is possible that the regression analysis may show no significant interaction even if one exists. Since interactions can be of considerable importance in medical studies, the preliminary inspection for interactions might be followed, where numbers permit, by a separate regression analysis for each differently-reacting subgroup, in order to describe the treatment effects as clearly as possible.

4.6 Summary

The analysis is considered of a set of data consisting of the times of death or censoring of 93 patients in a clinical trial, together with information on each patient about seven other factors which might influence survival. Relationships between the factors are investigated, and the data are broken down into subgroups and analysed by the methods of §2. Although numbers are small so that the breakdown cannot be complete enough for a satisfactory analysis, this is nevertheless a useful preliminary; a subsequent analysis using Cox's regression model is suggested, in which only two of the factors and their interaction need be considered.

References

Breslow, N. (1974), "Covariance analysis of censored survival data", Biometrics, 30, 89-99.

Cox, D.R. (1972), "Regression models and life tables", J. Roy. Statist. Soc. B, 34, 187-220.

Feigl, P. and Zelen, M. (1965), "Estimation of exponential survival probabilities with concomitant information", Biometrics, 21, 826-838.

Glasser, M. (1967), "Exponential survival with covariance", J. Amer. Statist. Assoc., 62, 561-568.

Kalbfleisch, J.D. and Prentice, R.L. (1973), "Marginal likelihoods based on Cox's regression and life-table model", Biometrika, 60, 267-278.

Kaplan, E.L. and Meier, P. (1958), "Nonparametric estimation from incomplete observations", J. Amer. Statist. Assoc., 53, 457-481.

Mantel, N. (1966), "Evaluation of survival data, and two new rank order statistics arising in its consideration", Cancer Chemotherapy Reports, 50, 163-170.

Mantel, N. and Haenszel, W. (1959), "Statistical aspects of the analysis of data from retrospective studies of disease", J. Nat. Cancer Inst., 22, 719-748.

Nelson, W. (1972), "Theory and application of hazard plotting for censored failure data", Technometrics 14, 945-966.

Peto, R. (1972), "Rank tests of maximal power against Lehmann-type alternatives", Biometrika 59, 472-475.

Peto, R. and Peto, J. (1972), "Asymptotically efficient rank invariant test procedures", J. Roy. Statist. Soc. A, 135, 185-206.

Peto, R. and Lee, P. (1973), "Weibull distributions for continuous-carcinogenesis experiments", Biometrics, 29, 457-470.

Roe, F.J.C., Peto, R., Kearns, F. and Bishop, D. (1970), "The mechanism of carcinogenesis by the neutral fraction of cigarette smoke condensate", Brit. J. Cancer, 24, 788-806.

Radhakrishna, S. (1965), "Combination of results from several 2×2 contingency tables", Biometrics, 21, 86-98.

Zelen, M. (1966), "Applications of exponential models to problems in cancer research", J. Roy. Statist. Soc. A, 129, 368-398.

Zippin, C. and Armitage, P. (1966), "Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter", Biometrics 22, 665-672.